



A I M S

LODE-BD Recommendations 2.0

*Report on how to select appropriate encoding strategies for producing **Linked Open Data (LOD)**-enabled **bibliographical data***

Prepared by:

Imma Subirats, FAO of the United Nations (Italy)

Marcia Lei Zeng, Kent State University (USA)

Abstract

LODE-BD aims to support the selection of appropriate encoding strategies for producing meaningful Linked Open Data (**LOD**)-enabled **bibliographical data** (directly or indirectly). The LODE-BD recommendations are applicable for structured data describing bibliographic resources such as articles, monographs, theses, conference papers, presentation materials, research reports, learning objects, etc. – in print or electronic format. The core component of **LODE-BD** contains a set of recommended decision trees for common properties used in describing a bibliographic resource instance. Each decision tree is delivered with various acting points and the matching encoding suggestions. The full range of options presented by **LODE-BD** will enable data providers to make their choices according to their development stages, internal data structures, and the reality of their practices.



Table of Contents

1. The LODE-BD Recommendations	3
1.1. Purpose of the LODE-BD Recommendations	3
1.2 The LODE-BD Report Roadmap	4
2. General Recommendations	5
2.1 Questions Addressed	5
2.2 Metadata Terms Overview	5
3. The Decision Trees: Recommendations for Individual Properties	8
3.1. Title Information	9
3.1.1. <i>Title/Alternative title</i>	9
3.2. Responsible Body	11
3.2.1 <i>Creator</i>	11
3.2.2. <i>Contributor</i>	13
3.2.3. <i>Publisher</i>	15
3.3. Physical Characteristics	17
3.3.1. <i>Date</i>	17
3.3.2. <i>Identifier</i>	19
3.3.3. <i>Language</i>	22
3.3.4. <i>Format / Medium</i>	23
3.3.5. <i>Edition/Version</i>	25
3.3.6. <i>Source</i>	27
3.4. Holding/Location Information	29
3.4.1. <i>Location / Availability</i>	29
3.5. Subject Information	31
3.5.1. <i>Subject</i>	31
3.6. Description of Content	33
3.6.1. <i>Description / Abstract / Table of Contents</i>	33
3.6.2. <i>Type/Form/Genre</i>	35
3.7. Intellectual property rights	36
3.7.1. <i>Right Statements</i>	36
3.8. Usage	38
3.8.1. <i>Audience / literary indication / education Level</i>	38
3.9. Relation	40
3.9.1. <i>Relation between resources</i>	40
3.9.2. <i>Relation between agents</i>	42
4. The step forward(With Further Readings)	44
4.1 Implementation Options	44
4.2 How to publish and consume Linked Data	45
4.3. Where to find Linked Data sets and Vocabularies	45
4.4. How to express metadata with different syntaxes: text, html, xml, rdf, and rdfa	46
4.5. Why publish bibliographic data as Linked Data?	46
Acknowledgement	46
Appendixes	47
Appendix 1. Background of the original report, LODE-BD 1.1	47
Appendix 2. Explanation of Terminology	48
Appendix 3. Metadata Standards used in LODE-BD	49
Appendix 4. Crosswalk of Metadata Term used in LODE-BD and schema.org terms	50

1. The LODE-BD Recommendations

With Web advances to an era of open and linked data, the traditional approach of sharing data within silos seems to have reached its end. From governments and international organizations to local cities and institutions, there is a widespread effort of opening up and interlinking their data. This report aims at providing bibliographic data providers of open repositories with a set of recommendations that will support the selection of appropriate encoding strategies for producing meaningful Linked Open Data (LOD)-enabled bibliographical data (LODE-BD).

Linked Data, a term coined by Tim Berners-Lee in his design note¹ regarding the Semantic Web architecture, refers to a set of best practices for publishing, sharing, and interlinking structured data on the Web. Key technologies that Linked Data builds on are: Uniform Resource Identifiers (URIs) for identifying entities or concepts in the world, RDF model for structuring and linking descriptions of things, HTTP for retrieving resources or descriptions of resources², and links to other related URIs in the exposed data to improve discovery of related information on the Web.

1.1. Purpose of the LODE-BD Recommendations

In the bibliographic universe there is a clear paradigm shift from fixed records to re-combinable metadata statements. For anyone who is contributing to an open bibliographic data repository as a data provider or service provider, the processes and strategies of providing data as Linked Data are practical issues. Guidelines and recommendations on what standards to follow and how to prepare LOD-ready metadata are essential.

There seems to be no one-size-fits-all approach because there existed a great number of metadata-related standards developed during the last two decades. They have been created by different communities for specific purposes to guide the design, creation, and implementation of data structures, data values, data contents, and data exchanges in certain communities. The operational metadata standards for data structures form a whole spectrum, ranging from independent ones (which do not reuse any metadata terms from a known namespace) to integrated ones (which would fully employ and incorporate existing metadata terms from other namespaces, usually seen in newly developed metadata application profiles and ontologies). Decisions regarding what standard(s) to adopt will directly impact the degree of LOD-readiness of the bibliographic data.

The approach of employing well-accepted metadata element sets and value vocabularies has already shown great benefits and potentials in terms of resource discovery, data reuse, data sharing, and the creation of new content based on Linked Data. However, deciding to take this approach is only the first step for the data providers and service providers of an open bibliographic repository. In the context of producing LOD-enabled bibliographical data, data and service providers are likely to have many specific questions related to the encoding strategies, for example:

- What metadata standard(s) should we follow in order to publish any bibliographic data as Linked Data?
- What is the minimal set of properties that a bibliographic dataset should include to insure meaningful data sharing?
- Is there any metadata model or application profile that can be directly adopted for producing bibliographical data (especially from our local database)?
- If the controlled vocabulary we have used is available as Linked Data, what kind of values should we exchange through our repository, specifically, the literal form representing a concept or the URI identifying the concept?
- How should we encode our data in order to move from a local database to a Linked Data dataset?

This report was born in this context with the purpose of assisting data providers in selecting appropriate encoding strategies for producing LOD-enabled bibliographical data (directly or indirectly). In order to enhance the quality of the interoperability and effectiveness of information exchange, the LODE-BD Recommendations are built on five key principles:

¹ Berners-Lee, Tim. 2007, Linked Data – Design Issues. <http://www.w3.org/DesignIssues/LinkedData> Last accessed: June 2012

² LOD2 Collaborative Project. 2010. Deliverable 12.5.1. Project fact sheet version 1.

http://static.lod2.eu/Deliverables/LOD2_D12.5.1_Project_Fact_Sheet_Version.pdf Last accessed: June 2012



1. To promote the use of well-established metadata standards and the emerging LOD-enabled vocabularies proposed in the Linked Data community;
2. To encourage the use of authority data, controlled vocabularies, and syntax encoding standards in metadata statements whenever possible;
3. To encourage the use of resource URIs as data values when they are available;
4. To facilitate the decision-making process regarding data encoding for the purpose of exchange and reuse;
5. To provide a reference support that is open for suggestions of new properties and metadata terms according to the needs of the Linked Data community.

1.2 The LODE-BD Report Roadmap

LODE-BD Recommendations are presented as a whole package, encompassing the important components that a data provider may encounter when deciding to produce sharable LOD-ready structured data describing **bibliographic resources** (such as articles, monographs, theses, conference papers, presentation material, research reports, learning objects, etc. – in print or electronic format) from a local database. In the future the recommendations may be extended to accommodate other kinds of information resources.

The recommendations are included in section 2 and 3 of this report:

- Section 2, **general recommendations**, presents nine groups of common properties identified by LODE-BD and the selected metadata terms to be used for describing bibliographic resources.
- Section 3, **decision trees**, demonstrates how to make decisions on selecting recommended properties according to the local needs.

Table 1. The Roadmap of the LODE-BD Report

Part	Focus	Document Explanation
1	About	Purpose of this report
2	General Recommendations	
2.1	Questions addressed	Three major questions addressed in the report
2.2	Metadata terms overview	All the metadata terms used in LODE-BD, presented in a crosswalk table
3	Decision Trees	A set of recommended decision-making trees for common properties used in describing a bibliographic resource instance. Each decision tree is delivered in a flowchart with various acting points. At the end of a decision tree a set of matching encoding suggestions is provided.
4	References	Recommended references for the decision-makers; Links to the general procedures of publishing Linked Data and useful syntax guidelines
5	Appendixes	Background information; The terminology used in the report; A list of metadata standards selected by LODE-BD

2. General Recommendations

2.1 Questions Addressed

Once a data provider has decided to publish a bibliographical database as Linked Data, there are important components that should be considered, including:

1. What kinds of entities and relationships are involved in describing and accessing bibliographic resources?

LODE-BD believes that a conceptual model would help to establish an overall picture of involving entities and relationships in bibliographic descriptions. In a broader context, the use of a similar conceptual model among data providers should also help foster a common understanding of the involving data models. Thus, LODE-BD uses a simple conceptual model based on three entities: *resource*, *agent* and *thema*. Major relations can be identified between a *resource* instance (e.g. an article or a report) and the *agent(s)* (e.g. a personal author or a research team) that are responsible for the creation of the content and the dissemination of the resource, as well as the *thema(s)* (e.g. things that being the subjects or topics of an article). The model provides sufficient capabilities for data providers to present their content (such as in document repositories and library catalogues) for sharing in the traditional environment or transferring to the Linked Data environment. (See explanations in a separate deliverable [Meaningful Bibliographic Metadata \(M2B\)](#)).

2. What properties should be considered for publishing meaningful/useful LOD-ready bibliographic data?

In the Linked Data context any data provider can expose anything contained in its local database. However, in the case of bibliographical data, standardized types of information should be considered in order to maximize the impact of sharing and connecting of the data. LODE-BD has built its recommendations on nine groups of common properties for describing bibliographic resources (details explained in [M2B](#)). These include specific best practice recommendations for about two-dozen properties used for describing a bibliographic resource as well as an additional two sets of properties for describing relations between bibliographic resources or between agents.

3. What metadata terms are appropriate in any given property when producing LOD-ready bibliographic data from a local database?

LODE-BD has selected a number of well-accepted and widely used metadata/vocabularies and used their metadata terms in the recommendations. All metadata terms used in the Recommendations are included in a crosswalk table (refer to Section 2.2). Flowcharts are used to present individualized **decision trees**, which provide adjustable decision process to data providers and for their situations when selecting metadata terms (refer to Section 3).

The comprehension of all the components below should enable a data provider to carry out the selection process of the metadata terms that fit in his bibliographic data requirements.

2.2 Metadata Terms Overview

All metadata terms corresponding to the properties grouped by LODE-BD are presented in the following crosswalk. Usually metadata terms from the Dublin Core namespaces are the fundamentals, while metadata terms from other namespaces are supplemented when additional needs are to be satisfied. They are:

- @prefix **dc**: <<http://purl.org/dc/elements/1.1/>> (Dublin Core Metadata Element Set namespace)
- @prefix **dcterms**: <<http://purl.org/dc/terms/>> (DCMI terms namespace)
- @prefix **bibo**: <<http://purl.org/ontology/bibo/>> (Bibliographic Ontology namespace)
- @prefix **agls**: <<http://www.agls.gov.au/agls/terms/>> (AGLS Metadata Standard namespace)
- @prefix **eprint**: <<http://purl.org/eprint/terms/>> (Eprints namespace)
- @prefix **marcrel**: <<http://id.loc.gov/vocabulary/relators/>> (MARC List for Relators namespace)

The semantics of the metadata terms (e.g. definition, usage, and relation with another property) defined by these specifications are inherited when a recommendation is made in a decision tree.

Table 2. Crosswalk of metadata terms used in the LODE-BD Recommendations

LODE-BD Group	Metadata Terms			
	General Metadata Terms		More Specific Metadata Terms	
	dc:-based	dcterms:-based		
1. Title Information	dc:title	dcterms:title	dcterms:alternative	
2. Responsible Body	dc:creator	dcterms:creator	bibo:editor	
	dc:contributor	dcterms:contributor	bibo:issuer	
	dc:publisher	dcterms:publisher	bibo:producer	
			bibo:distributor	
			bibo:owner	
3. Physical Characteristics	dc:date	dcterms:date	dcterms:created	
			dcterms:dateAccepted	
			dcterms:dateCopyrighted	
			dcterms:dateSubmitted	
			dcterms:modified	
			dcterms:valid	
			dcterms:available	
			dcterms:issued	
	dc:identifier	dcterms:identifier	bibo:asin	
			bibo:coden	
			bibo:doi	
			bibo:eanucc13	
			bibo:eissn	
			bibo:gtin14	
			bibo:handle	
			bibo:isbn	
			bibo:issn	
			bibo:lccn	
			bibo:oclcnum	
			bibo:pmid	
			bibo:sici	
			bibo:upc	
			bibo:uri	
	dc:language	dcterms:language		
	dc:format	dcterms:format	dcterms:medium	
	bibo:edition			
	bibo:status			
dc:source	dcterms:source	bibo:pages	bibo:pageStart	
			bibo:pageEnd	
		bibo:section		
		bibo:volume		
		bibo:issue		
		bibo:chapter		
4. Holding/Location Information	agls:availability	bibo:locator		

LODE-BD Group	Metadata Terms		
	General Metadata Terms		More Specific Metadata Terms
	dc:-based	dcterms:-based	
5. Subject Information	dc:subject	dcterms:subject	
	dc:coverage	dcterms:coverage	dcterms:spatial dcterms:temporal
6. Description of Content	dc:description	dcterms:description	dcterms:abstract dcterms:tableOfContent
	dc:type	dcterms:type	
7. Intellectual Property Rights	dc:rights	dcterms:rights	dcterms:rightsHolder dcterms:accessRights dcterms:license
8. Usage	dc:description	dcterms:description	
		dcterms:audience	dcterms:educationLevel dcterms:mediator
		dcterms:instructionalMethod	
9. Relation [between resources]	dc:relation	dcterms:relation	dcterms:isVersionOf
			dcterms:hasVersion
			dcterms:isReplacedBy
			dcterms:replaces
			dcterms:isRequiredBy
			dcterms:requires
			dcterms:isPartOf
			dcterms:hasPart
			dcterms:isReferencedBy
			dcterms:references
			bibo:translationOf
			bibo:annotates
[between agents]			bibo:citedBy
			bibo:cites
	eprint:affiliatedInstitution		
	eprint:grantNumber		
	marcrel:FND		

3. The Decision Trees: Recommendations for Individual Properties

To assist in the metadata term selection, this chapter provides **decision trees** for the properties included in each of the nine groups presented in the crosswalk table (refer to Section 2.2). Starting from the property that describes a resource instance, each flowchart presents decision points and gives a step-by-step solution to a given problem of metadata encoding. At the end of each flowchart, there are alternative sets of metadata terms for selection. A data provider can highlight the decision path and mark the metadata terms to be used at the end.

The types of values associated with a metadata term may be two types (see also Appendix B. Explanation of Terminology):

- literals (typically a string of characters; indicated by “string” in the flowcharts), or
- non-literals (a value which is a physical, digital or conceptual entity;³ indicated by “URI” in the flowcharts), depending on the requirements expressed in the namespace.

Text-based explanations corresponding to each of the flowcharts, with notes, steps, and examples, are also provided.

A flowchart is a diagrammatic representation that uses standardized symbols to portray steps and processes involved in decision making, with orders connected by flow lines with arrows. The basic shapes used in the figures follow the flowchart conventions:

Figure 1. Flowchart symbols and meanings

Name	Symbol	Used in flowchart
Narrow oval		Beginning of a decision tree
Flow-line		Direction of logic flow
Dashed Flow-line		Suggested direction of logic flow
Diamond		A decision to be made
Rectangle		A process to be carried out
Parallelogram		An input or available info. sources
Oval		End of a decision

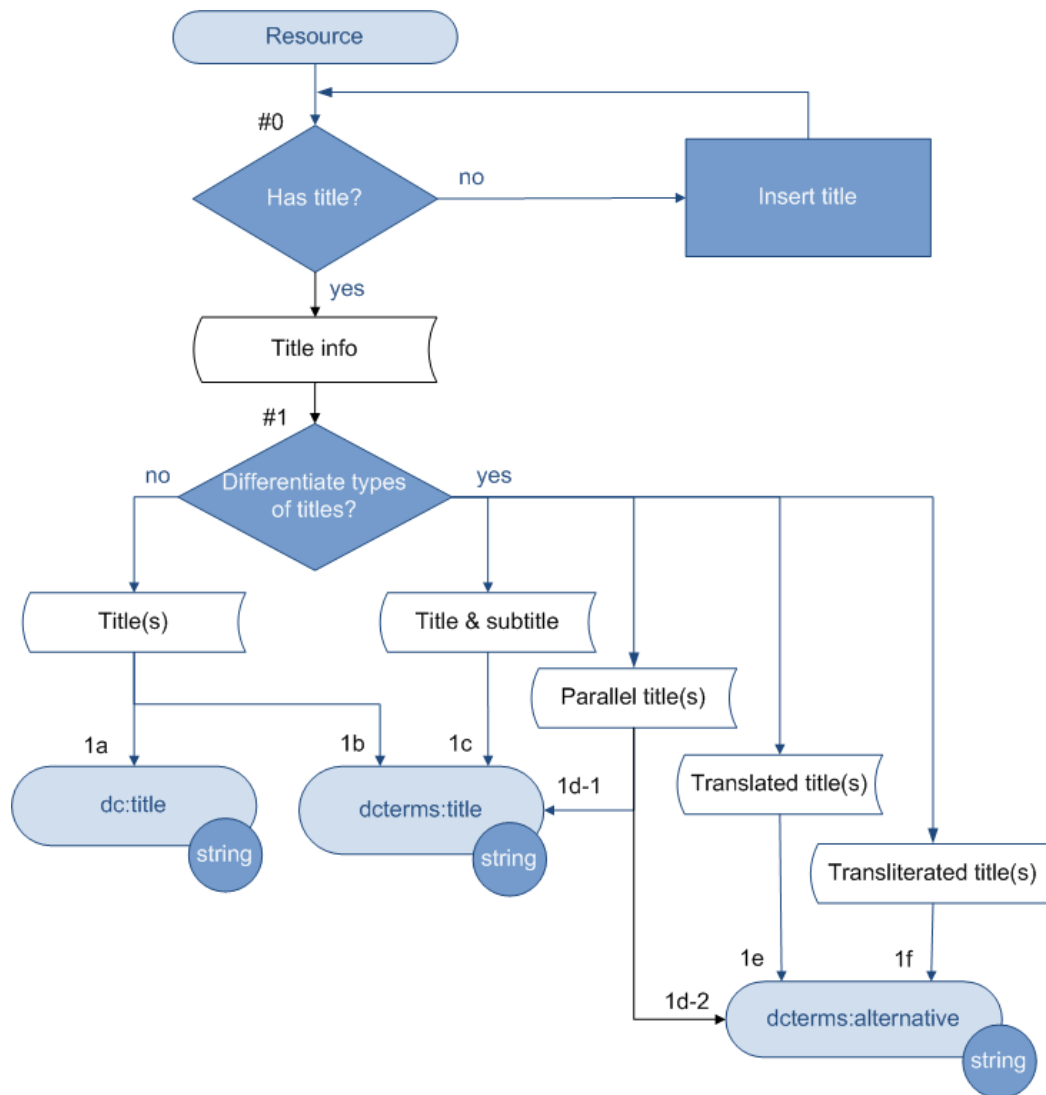
³ DCMI Abstract Model. <http://www.dublincore.org/documents/abstract-model/> Last accessed June 2012

3.1. Title Information

Title is one of the most important and relevant access points for any resource. The information is usually supplied through a number of properties including *title* and *alternative title* -- (handling subtitle(s), parallel title(s), translated title(s), transliterated title(s)). Title information is essential in the description of a resource; therefore the flowchart below foresees title as a mandatory metadata property.

3.1.1. Title/Alternative title

Relation with a resource being described: Resource has title.



Note

- Values for this property are always text strings.
- Although not emphasized in this report for the authority control of the titles of bibliographic resources given the context of this report, it is a logical step that resource titles, especially uniform titles, are also controlled.

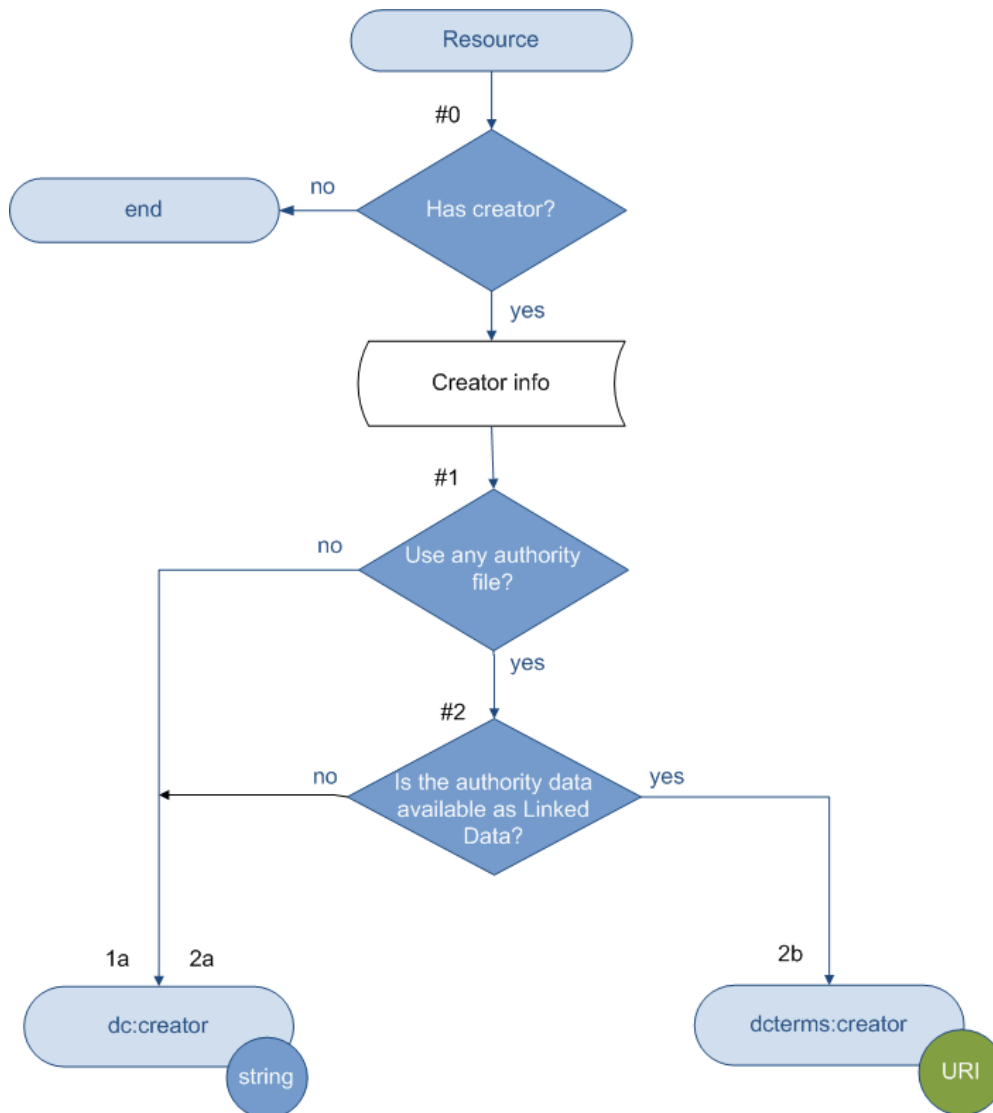
Decision	Question	Answer	Action	Value Type	Examples		
					Metadata Term	Value	
#0	Has title?	No	Insert title and go back to #0				
		Yes	Continue to #1				
#1	Differentiate types of titles?	No	1a	String	dc:title	Solar radiation energy and its utilization by Lucerne (Medicagosativa L.)	
			1b	String	dcterms:title	On the state of man [world agricultural situation]	
		Yes	title(s) and subtitle(s)	1c	String	dcterms:title	FAO yearbook of forest products 1996-2000
			parallel title(s)	1d-1	String	dcterms:title	Annuaire des produits forestiers de la FAO, 1996-2000
				1d-2	String	dcterms:alternative	
			translated title(s)	1e	String	dcterms:alternative	Anuario de productos forestales de la FAO, 1996-2000
							Working together for an International Alliance Against Hunger
			transliterated title(s)	1f	String	dcterms:alternative	Posly dobroj voli Prodovol'stvennoj i Sel'skokhozyajstvennoj Organizatsii Ob'edinennykh Natsij

3.2. Responsible Body

This group contains the properties associated with any agent who is responsible for the creation and/or publication of the content of the resource, for example, the *creator*, *contributor*, and *publisher* or *issuer* of a resource.

3.2.1 Creator

Relation with a resource being described: Resource has creator.



Note

- It is always recommended that an authority file be used for the responsible body that has created the resource.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has creator?	No	End			
		Yes	Continue to #1			
#1	Use any authority file?	No	1a	String	dc:creator	[Unauthorized form]: - Tim Berners-Lee - Tim B-L - Timothy John Berners-Lee - FAO of the UN - FAO Council (78th Session) Nov. 24, 1980, Rome, Italy
		Yes	Go to #2			
#2	Is the authority data available as Linked Data?	No	1b	String	dc:creator	[Authorized form]: - Berners-Lee, Tim - Food and Agriculture Organization of the United Nations - FAO Council (Sess. 78 : 24 Nov 1980 : Rome, Italy)
		Yes	2b	URI	dcterms:creator	http://aims.fao.org/aos/corporate/c_1297 [1] http://aims.fao.org/aos/conference/c_1842 [2] http://viaf.org/viaf/85312226/#Berners-Lee, Tim [3] http://www.w3.org/People/Berners-Lee/card [4] [URI of a responsible body]

[1] A corporate body's URI, from the FAO Authority Description Concept Scheme

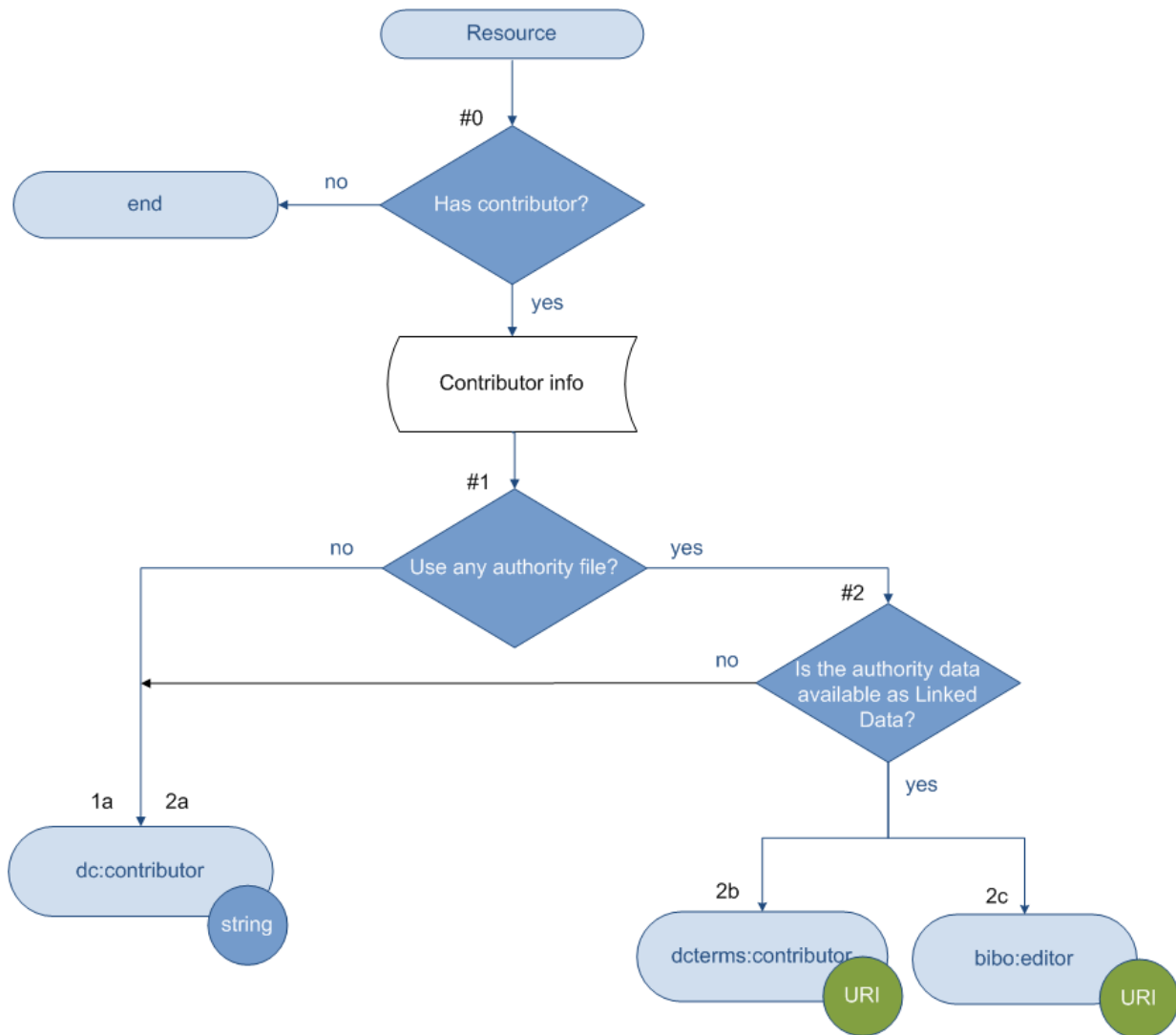
[2] A conference' URI, from the FAO Authority Description Concept Scheme

[3] Tim Berners-Lee's URI, from the VIAF (Virtual International Authority File)

[4] Tim Berners-Lee's URI: <http://www.w3.org/People/Berners-Lee/card#i> (Source of note: <http://www.linkedin.com/in/timbl>)

3.2.2. Contributor

Relation with a resource being described: Resource has contributor.



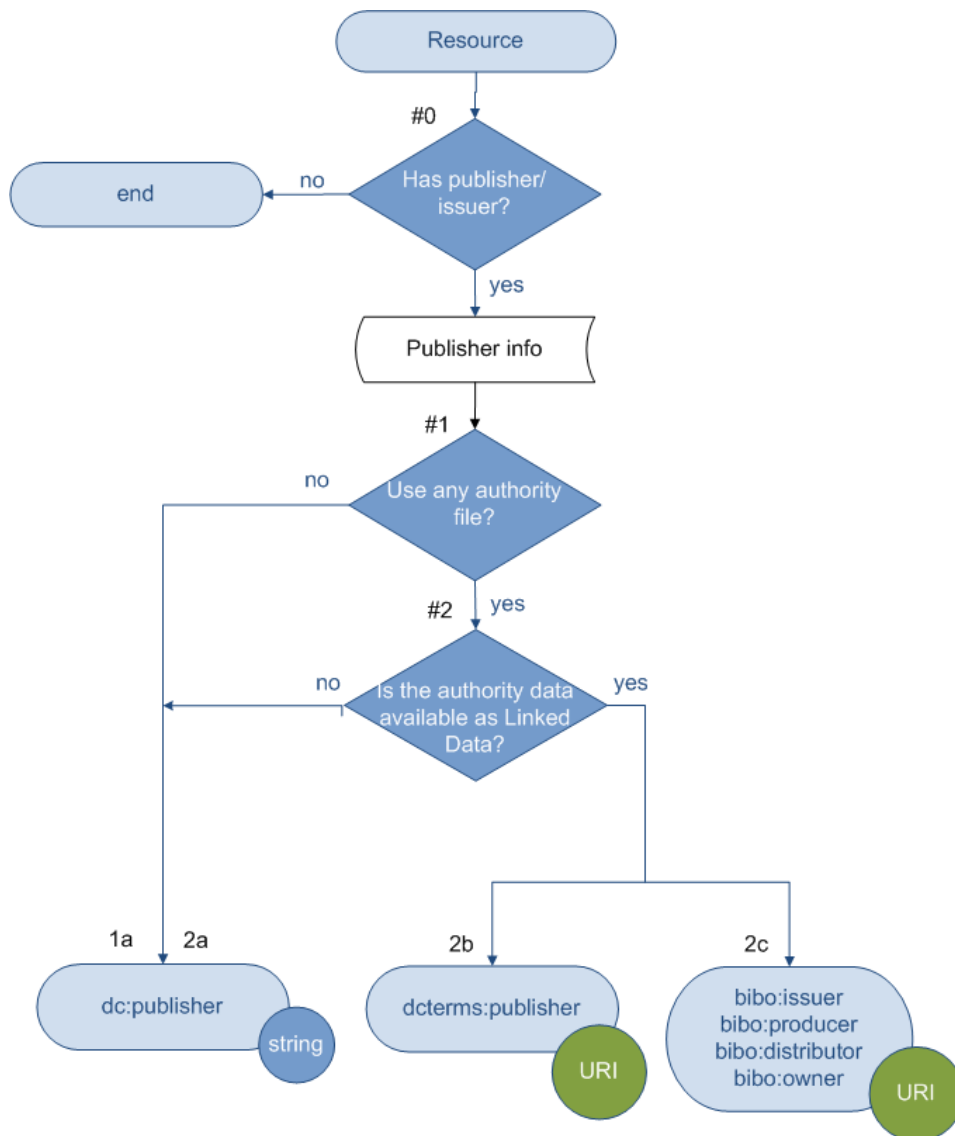
Note

- It is always recommended that an authority file be used for a responsible body that has contributed to the resource.

Decision	Question	Answer	Action	Value Type	Metadata Term	Examples Value
#0	Has contributor?	No	End			
		Yes	Continue to #1			
#1	Use any authority file?	No	1a	String	dc:contributor	[Unauthorized form]: - Tim Berners-Lee - Tim B-L - Timothy John Berners-Lee - FAO of the UN - FAO Council (78th Session) - Nov. 24, 1980, Rome, Italy
		Yes	Go to #2			
#2	Is the authority data available as Linked Data?	No	2a	String	dc:contributor	[Authorized form]: - Berners-Lee, Tim - Food and Agriculture Organization of the United Nations
			2b	URI	dcterms:contributor	[URI of a responsible body]
			2c	URI	bibo:editor	

3.2.3. Publisher

Relation with a resource being described: Resource has publisher.



Note

- It is always recommended that an authority file be used for a responsible body that is responsible for publishing or producing the resource.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has publisher/issuer?	No	End			
		Yes	Continue to #1			
#1	Use any authority file?	No	1a	String	dc:publisher bibo:owner	[Un-authorized form] : • FAO Rome (Italy) • FAO • F.A.O. • FAO of the UN • FAO, Rome • Food and Agriculture Organization • F.A.O. of the U.N. [Authorized form] : • Food and Agriculture Organization of the United Nations
		Yes	Go to #2			
#2	Is the authority data available as Linked Data?	No	2a (See#1a)			
			Use authorized form from an authority file]			
		Yes	2b	URI	dcterms: publisher	[URI of a responsible body]
			2c	URI	bibo: issuer bibo:producer bibo:distributor bibo:owner	[URI of a responsible body]

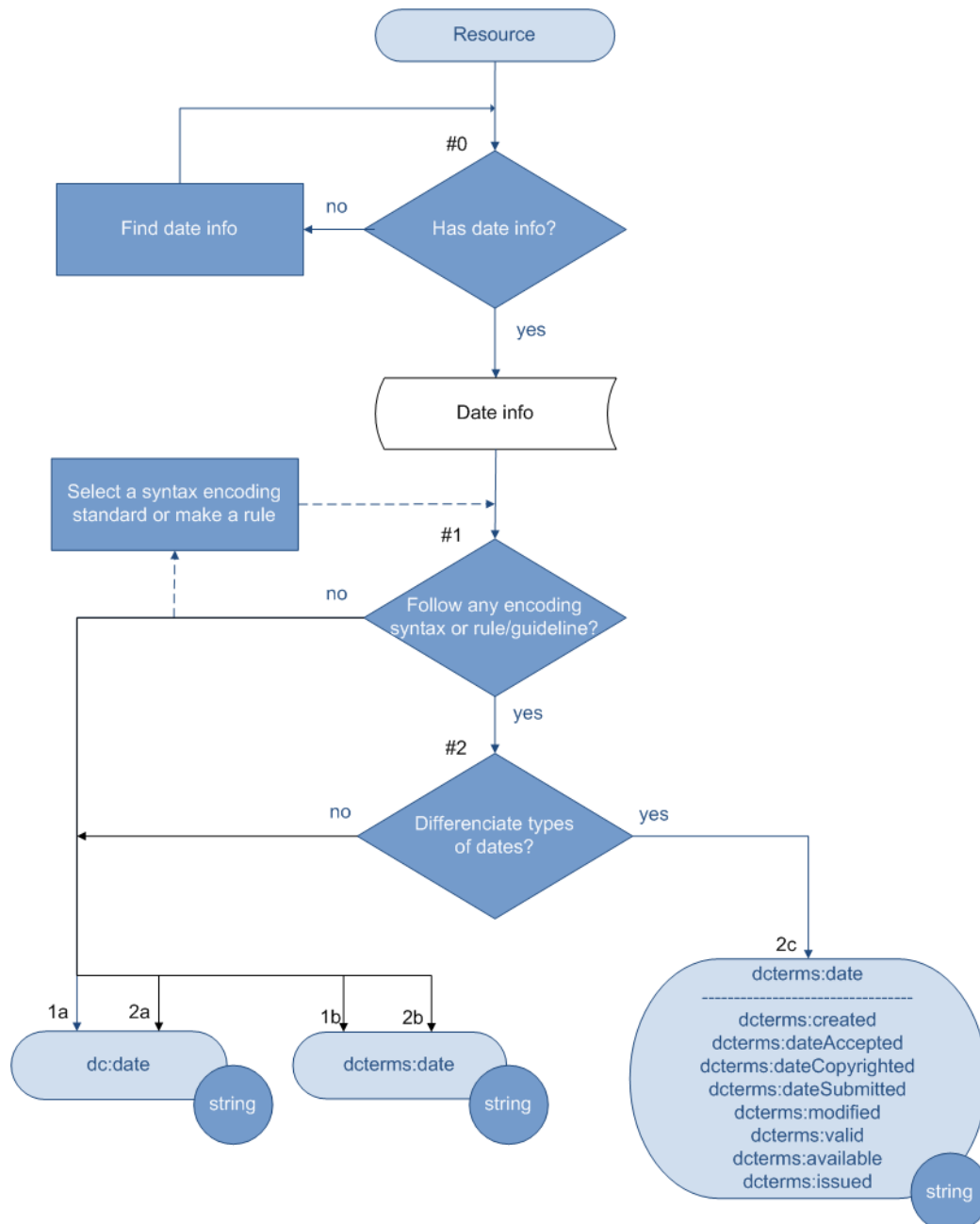
3.3. Physical Characteristics

Properties that describe the appearance and the characteristics of the physical form of a resource are placed into this group. They are: *date*, *identifier*, *language*, *format/medium*, *edition/version*, and *source*.

3.3.1. Date

Relation with a resource being described: Resource has date.

Date is considered essential information in the description of a resource; therefore the flowchart below foresees **date** as a mandatory property.



Note

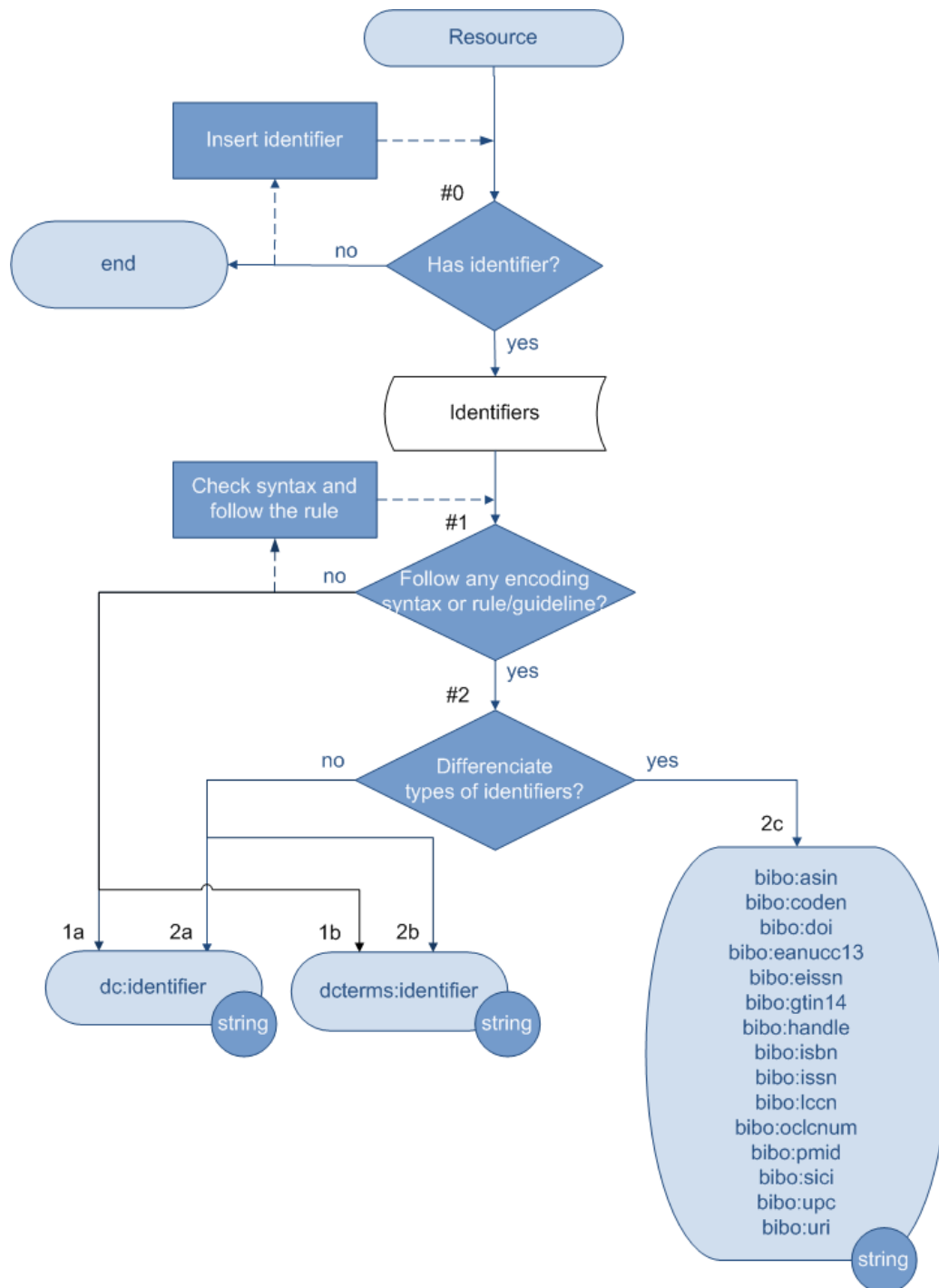
- Recommended best practice is to follow an encoding syntax, such as that defined by the W3CDTF profile of ISO 8601.

Decision	Question	Answer	Action	Value Type	Examples		
					Metadata Term	Value	
#0	Has date?	No	Find date info and go back to #0				
		Yes	Continue to #1				
#1	Follow any encoding syntax or rule/guideline?	Yes	Continue to #2				
		No	1a	String	dc:date	[198?] [1996] [1997?] 1968-2006 7 Jul 1989 7 July 1989 7-July-1989 Jul 1989 1989 Jul Jan-Feb 1997 1-5 Feb 1997 Spr 1997 20 Mar - 15 Apr 1995	
			1b	String	dcterms:date	[see all examples above]	
			2a	String	dc:date	1997	
#2	Differentiate type of dates?	No	2b	String	dcterms:date	1997-07	
			Yes	2c	String	dcterms:date	1997-07-16
						dcterms:created	1997-07-16T19:20+01:00
		dcterms:dateAccepted				1997-07-16T19:20:30+01:00	
		dcterms:dateCopyrighted				1997-07-	
		dcterms:dateSubmitted				16T19:20:30.45+01:00 [1]	
		dcterms:modified					
		dcterms:valid					
		dcterms:available					
dcterms:issued							

[1] W3C. (1997) Date and Time Formats. <http://www.w3.org/TR/NOTE-datetime>

3.3.2. Identifier

Relation with a resource being described: Resource has identifier.



Note

- It is always recommended that a resource has an identifier or identifiers.
- Established codes for identifiers (universal or local) should be used for any kind of identifiers. It is always recommended to check the syntax, follow or create a rule/guideline when handling identifiers.
- In the bibliographic descriptions, a resource is always represented by a unique ID. This ID may be locally assigned (or temporarily being local) [1], or be the same as its global recognizable identifiers such as a URI [2], or contains the string that is from a universal identifier such as an ISSN or a DOI [3].

Examples (from Bibliographic Ontology (bibo:))

[1] <info:doi/10.1134/S0003683806040089> a bibo:Article

[2] <http://www.nytimes.com/2007/07/09/us/09cnd-penn.html> a bibo:Article;

<http://www.amazon.com/dp/026256212X"> a bibo:Document

<urn:isbn:23983498> a bibo:Book

<urn:issn:23346587> a bibo:Journal

<http://www.w3.org/2007/Talks/0619-Nancy-IH/> a cc:Work, bibo:Slideshow

<http://ic2008.loria.fr/> a bibo:Conference

[3] <http://www.zotero.org/services/urn/isbn/026256212X"> a bibo:Book

In this report, such a unique ID is assumed to each resource being described, at the beginning of a decision tree.

- In addition to this unique ID, there are identifiers that are assigned to the original resource within the domains of various systems such as ISBN, DOI, ISSN, etc. The decision tree presented here is about **those** identifiers, even though one of the identifiers is the same as the unique ID of the resource being described.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has identifier?	No	End but recommended to insert an identifier			
		Yes	Continue to #1			
#1	Follow any encoding syntax, /rule /guideline?	No*	1a	String	dc:identifier	http://www.ukoln.ac.uk/ [1] urn:ietf:rfc:1766 [1]
			1b	String	dcterms:identifier	http://www.ukoln.ac.uk/ [1] urn:ietf:rfc:1766 [1]
		Yes	Continue to #2			
		#2	Differentiate types of identifiers?	No	2a	String
2b	String				dcterms:identifier	http://www.ukoln.ac.uk/ [1] urn:ietf:rfc:1766 [1]
Yes	2c			String	bibo:asin	020530902X [2]
					bibo:coden	66HYAL [3]
					bibo:doi	doi:10.1109/ISSTA.2002.1048560 [4]
					bibo:eanucc13	0123456789012 [5]
					bibo:eissn	0378-5955 [6]
					bibo:gtin14	00012345600012 [7]
					bibo:handle	http://hdl.handle.net/10760/6634 [8]
					bibo:isbn	9-788175-257665 [9] 9788175257665
					bibo:issn	0317-8471 [10]
					bibo:lccn	79051955 [11]
					bibo:oclcnum	ocm00012345 [12] ocn123456789
					bibo:pmid	20346624 [13]
					bibo:sici	0095-4403(199502/03)21:3 <12:WATIIB>2.0.TX;2-J [14]
					bibo:upc	5778400002 [15]
					bibo:uri	http://example.org/absolute/URI/with/ absolute/path/to/resource.txt [16] ftp://example.org/resource.txt

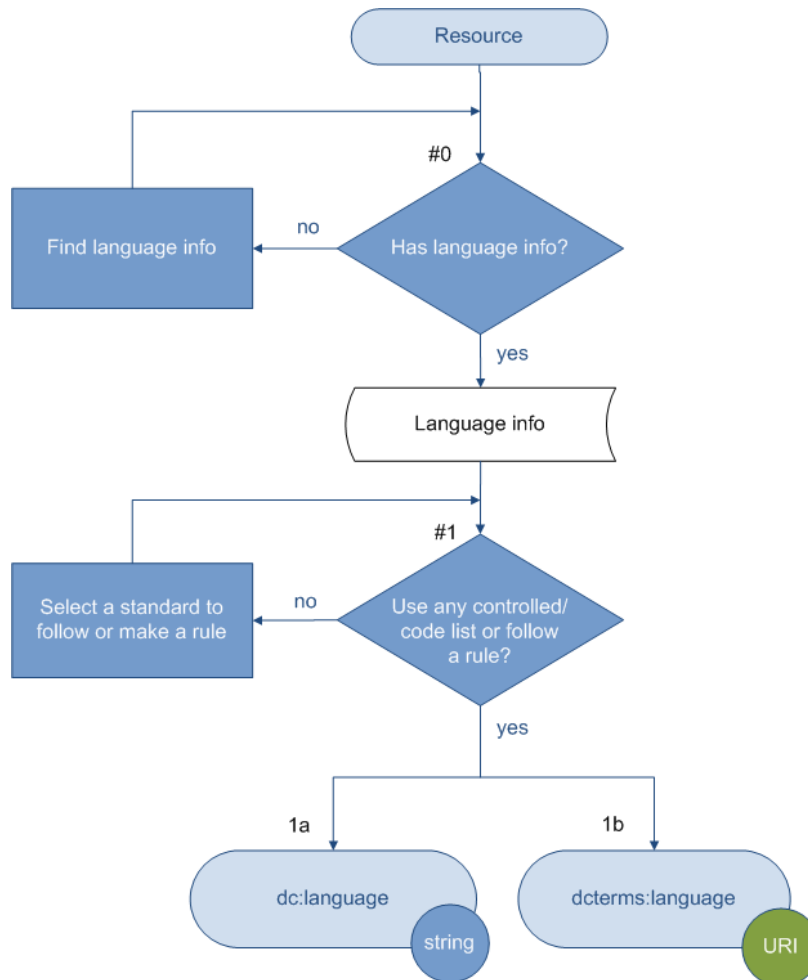


- [1] From <http://dublincore.org/documents/2001/04/12/usageguide/simple-html.shtml>
- [2] From http://en.wikipedia.org/wiki/Amazon_Standard_Identification_Number
- [3] From <http://en.wikipedia.org/wiki/CODEN>
- [4] From <http://www.doi.org/>
- [5] From <http://www.gtin.info/>
- [6] From <http://en.wikipedia.org/wiki/EISSN>
- [7] From <http://www.gtin.info/>
- [8] From <http://eprints.rclis.org/>
- [9] From http://en.wikipedia.org/wiki/International_Standard_Book_Number
- [10] From http://en.wikipedia.org/wiki/International_Standard_Serial_Number
- [11] From <http://catalog.loc.gov/>
- [12] From <http://www.oclc.org/batchprocessing/controlnumber.htm>
- [13] From <http://www.ncbi.nlm.nih.gov/pubmed/>
- [14] From http://en.wikipedia.org/wiki/Serial_Item_and_Contribution_Identifier
- [15] From http://en.wikipedia.org/wiki/Universal_Product_Code
- [16] From http://en.wikipedia.org/wiki/Uniform_Resource_Identifier

3.3.3. Language

Relation with a resource being described: Resource has language information.

Language is considered essential information in the description of a resource; therefore the flowchart below foresees **language** as a mandatory property.



Note

- Recommended best practice is to use an encoding scheme, such as the three-letter code (ISO639-2) or the two-letter code (ISO639-1).

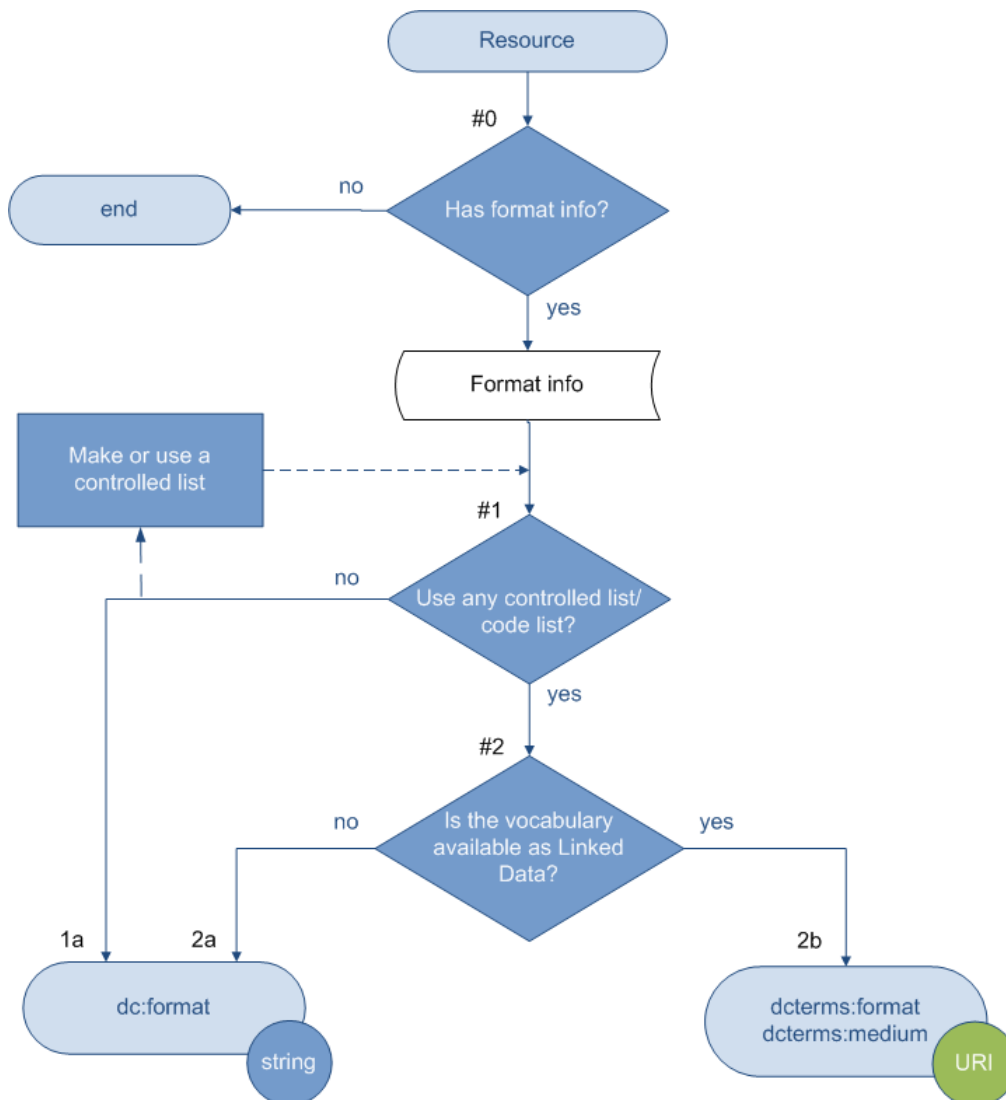
Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has language info?	No	Find language info and go back to #0			
		Yes	Continue to #1			
#1	Use any controlled list /code list or follow a rule?	No	Go back to #1			
		Yes	1a	String	dc:language	cat [1] ca [2]
			1b	URI	dcterms:language	[URI of a language name]

[1] From ISO639-2 http://www.loc.gov/standards/iso639-2/php/code_list.php

[2] From ISO639-1

3.3.4. Format / Medium

Relation with a resource being described: Resource has format.



Note

- It is always recommended that a controlled vocabulary be used for your collection when describing 'format', such as the list of Internet Media Types [MIME].
- It is also recommended that a controlled vocabulary be used for your collection when using dcterms:medium. Because dcterms:medium has the definition of material or physical carrier of the resource, the Internet Media Types [MIME] should NOT be used for these values.
- If no formal controlled vocabulary exists, handle the media type like another resource.[1]

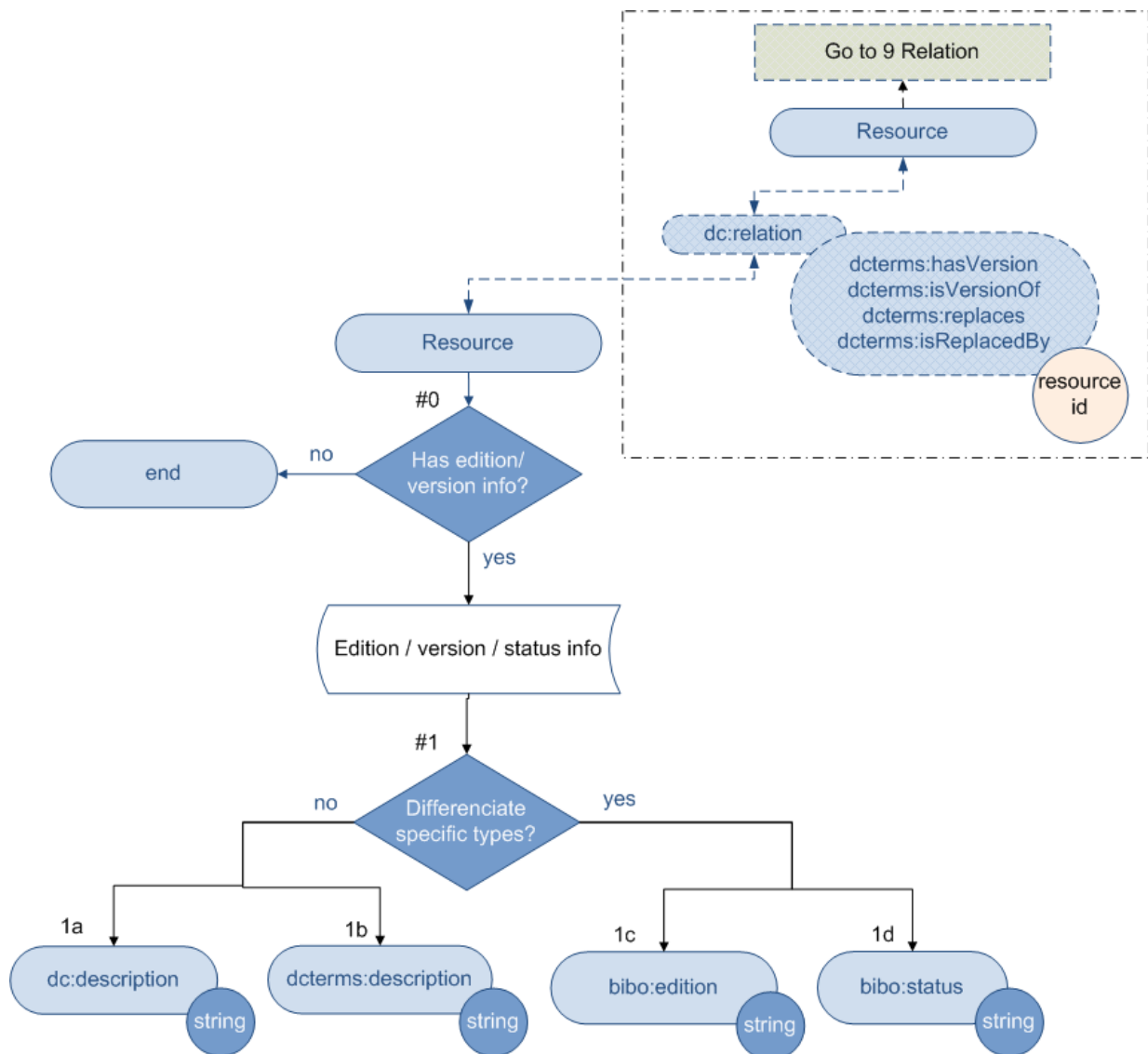
Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has format info?	No	End			
		Yes	Continue to #1			
#1	Use any controlled list or code list?	No	1a	String	dc:format	html
		Yes	Continue to #2			
#2	Is the controlled vocabulary available as Linked Data?	No	2a	String	dc:format	text/html
		Yes	2b	URI	dcterms:format	mime:jpeg
					dcterms:medium	:oilOnWood [1]

[1] See guidelines and example at:

http://wiki.dublincore.org/index.php/User_Guide/Creating_Metadata#Guidelines_for_the_creation_of_medium_content

3.3.5. Edition/Version

Relation with a resource being described: Resource has edition/version/status.



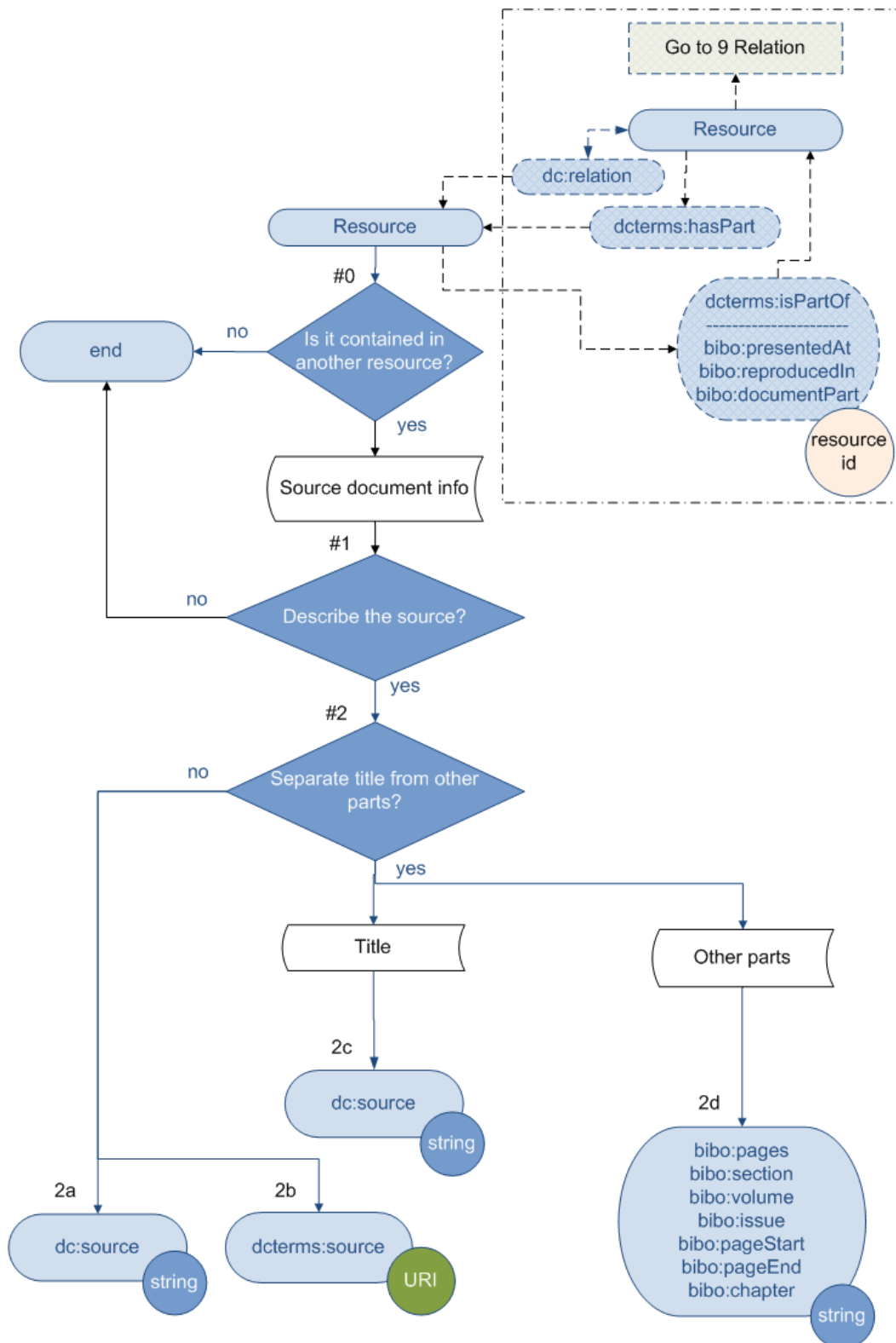
Notes

- When an edition or version of a resource is to be described, the relation between a resource and its related version(s) should also be described. In this graph, a dash-lined box signifies such relation(s) and points to Section 9, “Relation”, in this report.
- The main body of the graph only focuses on the description of edition or version as a part of the physical characteristics of a resource. For describing relations between different versions of resources, go to Section 9.1 Relations between resources.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has edition version info?	No	End			
		Yes	Continue to #1			
#1	Differentiate specific types?	No	1a	String	dc:description	2 nd ed.
			1b	String	dcterms:description	2 nd ed.
		Yes	1c	String	bibo:edition	2 nd ed.
			1d	String	bibo:status	Final

3.3.6. Source

Relation with a resource being described: Resource has source.



Notes

- When a resource to be described is contained in another resource, the relations between the resources may be described according to the convention of a data provider. In this graph, a dash-lined box signifies such relation(s) and points to Section 9, "Relation".
- The main body of the graph is only focused on the description of the source of a resource. For describing relations between the resources involved, go to Section 9.1 Relation between resources.
- It is recommended that if the resource titles are controlled through an authority file, use the controlled title or identifier.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Is it contained in another resource?	No	End			
		Yes	Continue to #1			
#1	Describe the resource?	No	End			
		Yes	Continue to #2			
#2	Separate title from other parts?	No	2a	String	dc:source	Proceedings of the National Academy of Sciences of the United States of America, v. 95(10) p. 5632-5636
						http://www.pnas.org/content/by/year/2010
			2b	URI	dcterms:source	http://www.pnas.org/content/by/year/2010
		Yes	2c	String (Title)	dc:source	Proceedings of the National Academy of Sciences of the United States of America
			2d	String (Other parts)	bibo:pages	542
					bibo:section	2
					bibo:volume	95
					bibo:issue	10
					bibo:pageStart	5632
					bibo:pageEnd	5636
					bibo:chapter	II

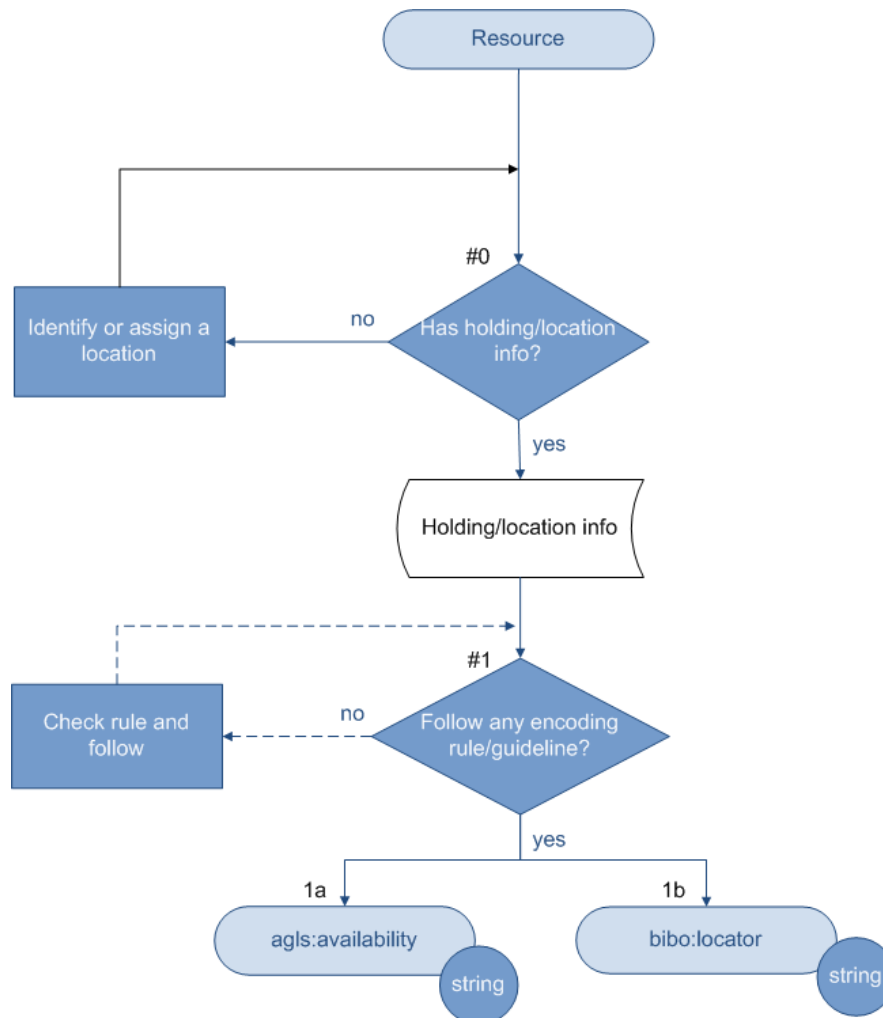
3.4. Holding/Location Information

It is important for a resource to be located and obtained in the information exchange. Properties that record the *location* and *availability* information are taken into account in this unique group.

3.4.1. Location / Availability

Relation with a resource being described: Resource has holding or location information.

Location is considered essential information in the description of a resource in a digital repository; therefore the flowchart below foresees holding or location information as a mandatory property.



Note

- It is always recommended that location information be provided consistently by following an encoding rule or guideline.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has holding/location info?	No	Identify or assign a location and Go back to #0			
		Yes	Continue to #1			
#1	Follow any encoding rule or guideline?	No	Go back to #1			
		Yes	1a	String	agls:availability	http://www.example.org/services/id5678/ Contact the Publications Section on 1300 999 999[1] University of Vienna, Peter Jordanstr. 52, A-1190 Vienna, Austria
			1b	String	bibo:locator	Box 12, Folder 3

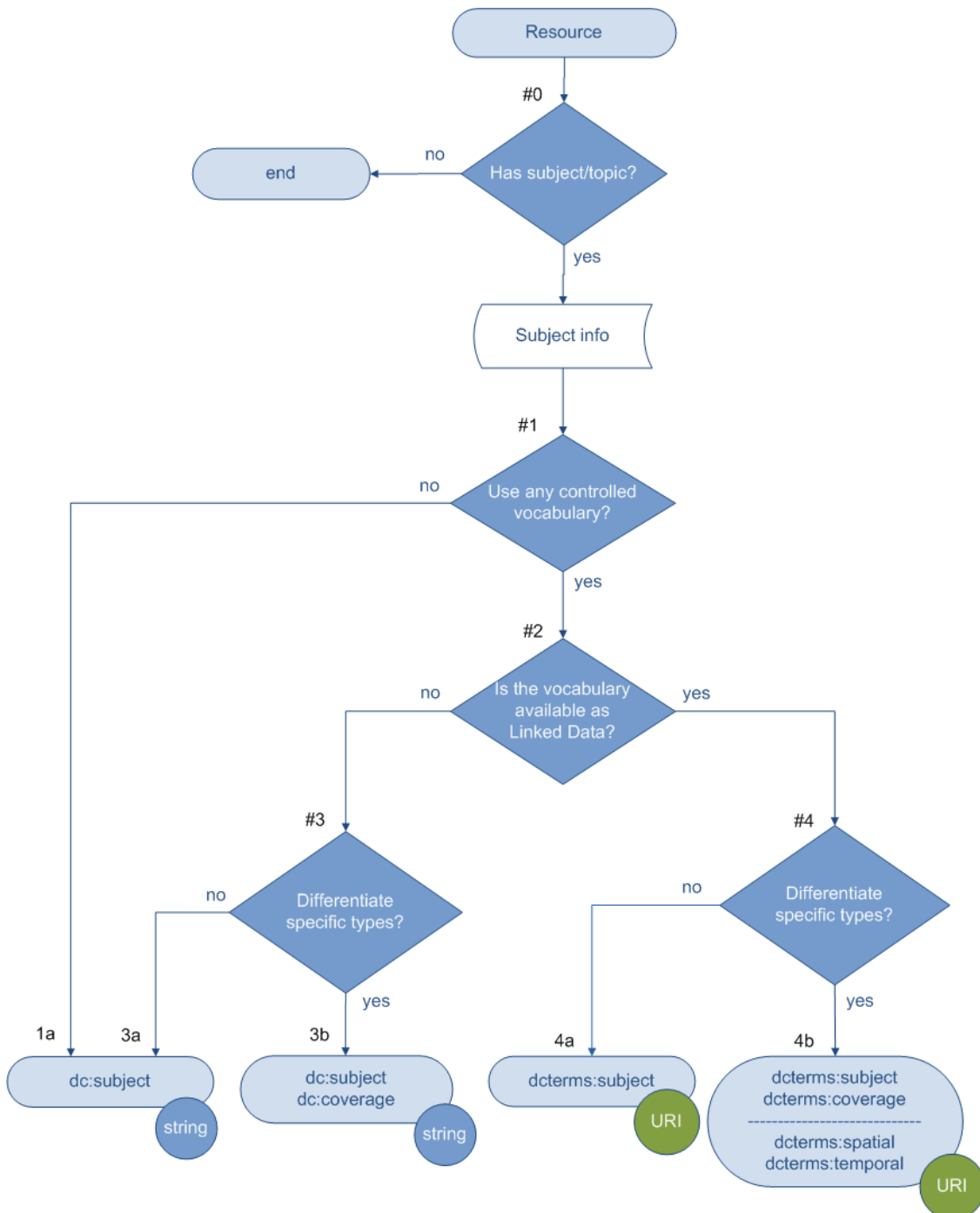
[1] From <http://www.agls.gov.au/>

3.5. Subject Information

In contrast to the physical characteristics, the Subject group embraces the properties that describe or otherwise help the identification of what the resource is about or denotes, in the form of *subject term*, *classification/category*, freely assigned *keyword*, *geographic term*, and so on.

3.5.1. Subject

Relation with a resource being described: Resource has subject/topic.



Note

- It is always recommended to index the concept/topic/subject/category of a resource. Examples of values include: concepts represented by terms from a controlled vocabulary; keywords; classes or categories represented by notations or labels from a classification system.
- More and more controlled vocabularies are published as Linked Data where concepts are represented by non-literal values (i.e., an identifier and/or a http URI). For example, each AGROVOC concept has its unique http URI. LODE-BD recommends using these URIs instead of the literal forms (i.e., the labels) as values when considering moving towards publishing your data as Linked Data.

Examples of values:

- a concept URI of your own controlled vocabulary;
 - a URI of a concept from a published thesaurus (e.g., EuroVoc) or classification (e.g., Dewey Decimal Classification);
 - a URI of an agent when the agent is the subject/focus of a resource (e.g., URI of a conference defined in a foaf file, a URI of a person from VIAF).
- Usually a value encoding scheme's title (e.g., AGROVOC or LCSH) should be indicated along with the value. Also when using literal forms than URIs, the language of the words should be indicated. (Consult references in the appendix if needed).

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has subject/topics?	No	End			
		Yes	Continue to #1			
#1	Use any controlled vocabulary?	No	1a	String	dc:subject	paddy Pacific Islands & Oceania 19th century
		Yes	Continue to #2			
#2	Is the vocabulary available as linked data?	No	Continue to #3			
		Yes	Continue to #4			
#3	Differentiate types of subjects?	No	3a	String	dc:subject	Rice Pacific Islands Nineteenth century
			3b	String	dc:subject dc:coverage	Rice Pacific Islands Nineteenth century
		Yes	4a	URI	dcterms:subject	http://aims.fao.org/aos/agrovoc/c_6599 [2] http://aims.fao.org/aos/agrovoc/c_5487 [3]
			4b	URI	dcterms:subject dcterms:coverage dcterms:spatial dcterms:temporal	http://aims.fao.org/aos/agrovoc/c_6599 [2] http://aims.fao.org/aos/agrovoc/c_5487 [3] http://aims.fao.org/aos/agrovoc/c_5487 [3] http://id.loc.gov/authorities/sh85091984 [4]

[1] From the Dewey Decimal Classification: "586 Seedless plants" (English version).

[2] http://aims.fao.org/aos/agrovoc/c_6599 is the URI of a concept in AGROVOC. Its preferred English label is "Rice".

[3] http://aims.fao.org/aos/agrovoc/c_5487 is the URI of a concept in AGROVOC. Its preferred English label is "Pacific Islands".

[4] <http://id.loc.gov/authorities/sh85091984> is the URI of a concept in LCSH. Its preferred English label is "Nineteenth century".

[5] The URI of the Dewey Decimal Classification: "586". Its English caption is "Seedless plants".

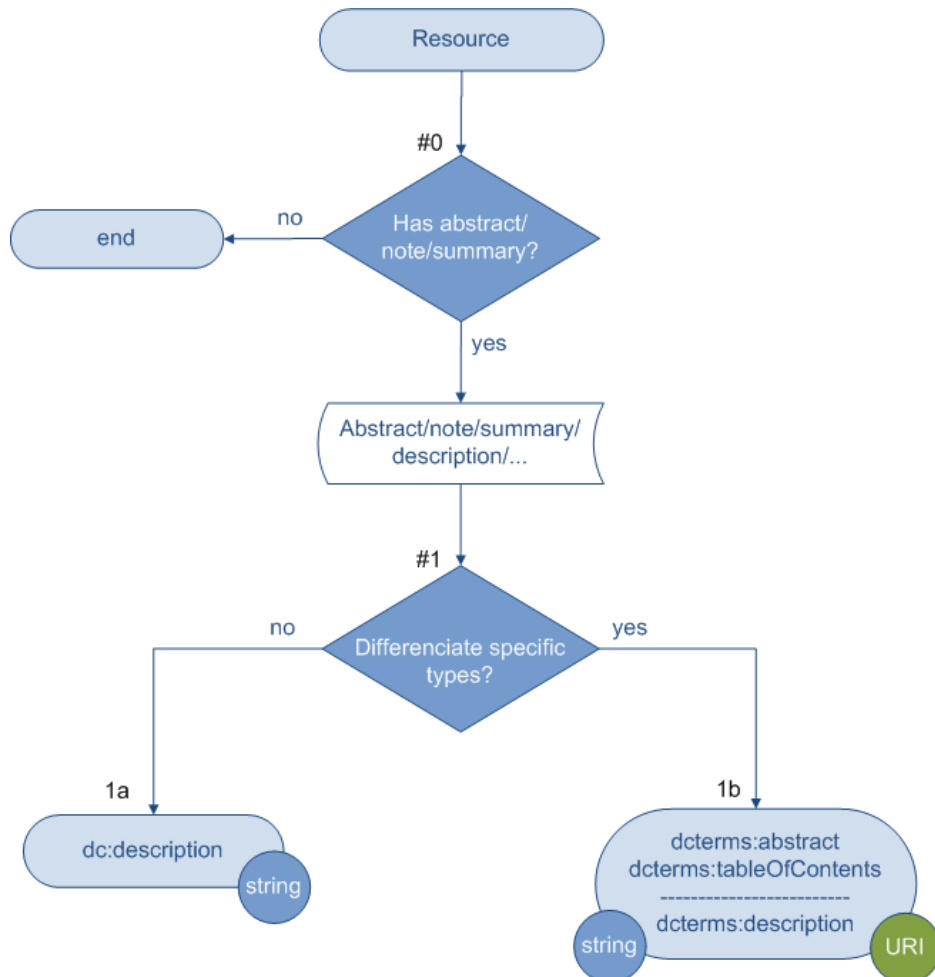


3.6. Description of Content

Two major types of descriptions that focus on the content of the resource rather than the physical object are considered in this group: a) any representative description of the content, usually in the form of *abstract*, *summary*, *note*, and *table of contents*; and b) *type* or *genre* of the resource.

3.6.1. Description / Abstract / Table of Contents

Relation with a resource being described: Resource has description, abstract or table of contents.



Note

- In describing the content, different words might have been used, such as “abstract” vs. “note”, or “description” vs. “summary”. A table of contents may also be presented in a description.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has abstract/note/summary?	No	End			
		Yes	Continue to #1			
#1	Differentiate types of content descriptions?	No	1a	String	dc:description	One of the least understood aspects of population biology is ...
		Yes	1b	String or URI	dcterms:abstract	One of the least understood aspects of population biology is ... http://jeclap.oxfordjournals.org/content/2/4/391.abstract [1]
					dcterms:table-of-contents	Introduction -- Formal theory -- Coevolution -- http://www.library.cornell.edu/preservation/tutorial/toc.html [2]
					dcterms:description	Contains a series of articles which are intended to ... VocBench is a web-based, multilingual, vocabulary editing and workflow tool developed by FAO. It ... [3] http://aims.fao.org/tools/vocbench-2 [3]

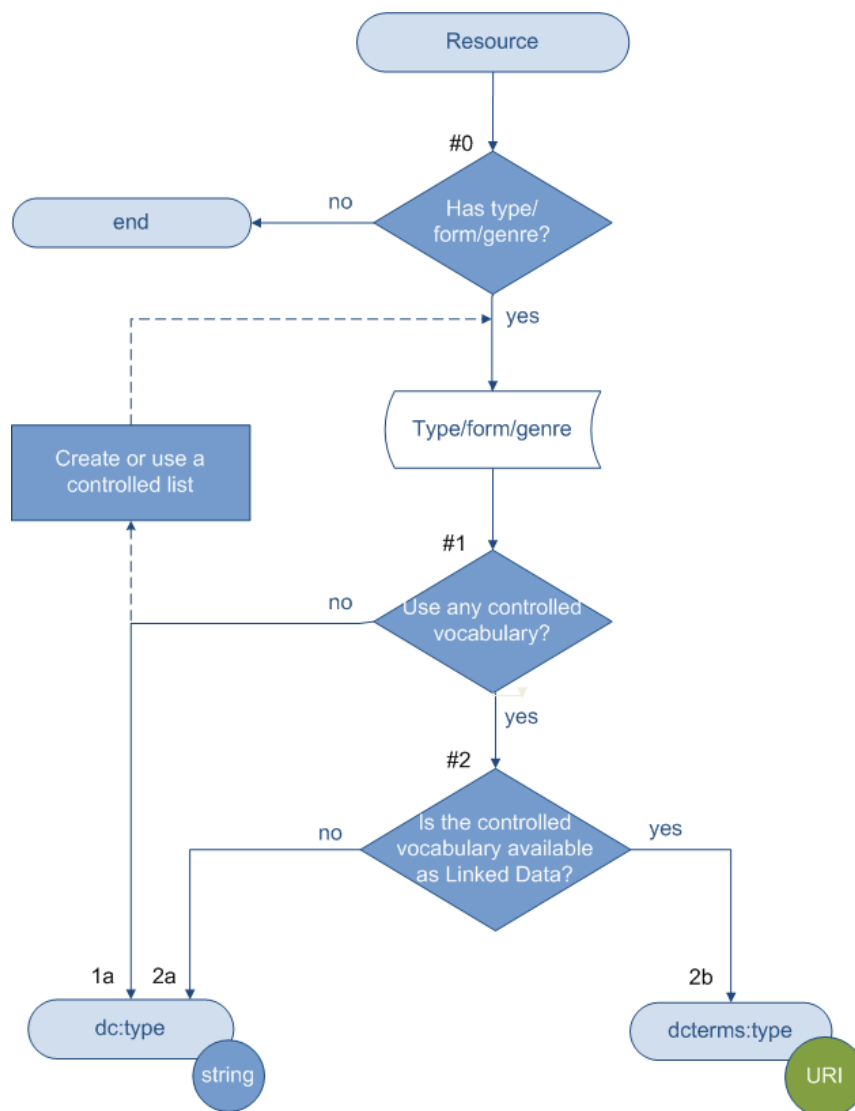
[1] The URL is the abstract of a report “Transatlantic Airline Alliances: The Joint EU–US Report” published by *Journal of European Competition Law & Practice* (2011) 2 (4).

[2] The URL is the Table of Contents page of [Moving Theory into Practice: Digital Imaging Tutorial](http://www.library.cornell.edu/preservation/tutorial/toc.html), Cornell University Library/Research Department, 2000-2003.

[3] Both the text and URL are from the VocBench Webpage, FAO of the United Nation.

3.6.2. Type/Form/Genre

Relation with a resource being described: Resource has type/form/genre.



Note

- It is always recommended that a controlled vocabulary be used or created for your collection when describing a resource type.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has type/form/genre?	No	End			
		Yes	Continue to #1			
#1	Use any controlled vocabulary ?	No	1a	String	dc:type	Lecture; Poster, ...
		Yes	Continue to #2			
#2	Is the controlled vocabulary available as linked data?	No	2a	String	dc:type	Interactive Resource
		Yes	2b	URI	dcterms:type	http://purl.org/dc/dcmitype/InteractiveResource [1]

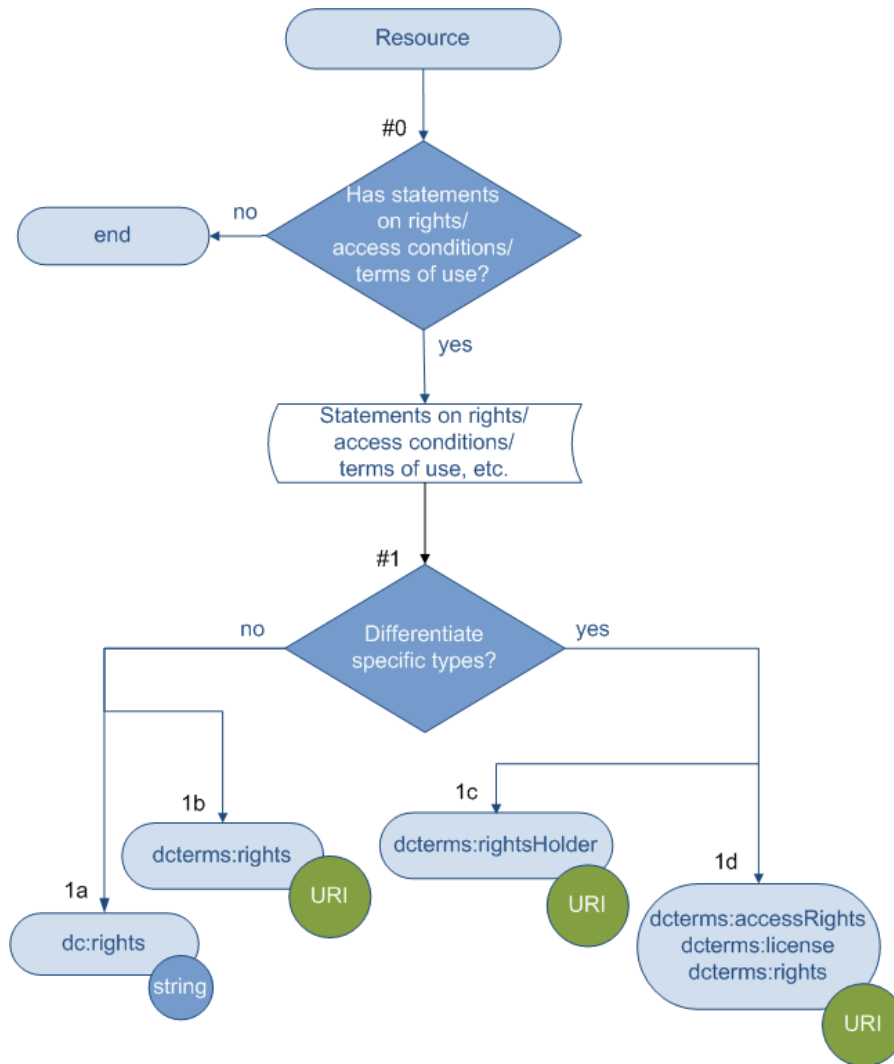
[1] <http://purl.org/dc/dcmitype/InteractiveResource> is the URI of the concept "Interactive Resource", from *DCMI Type Vocabulary*.

3.7. Intellectual property rights

Any property that deals with an aspect of intellectual property rights relating to access and use of a resource is included in this group, with special regard to *rights*, *terms of use*, and *access condition*.

3.7.1. Right Statements

Relation with a resource being described: Resource has intellectual property rights statement.



Note

- The property may be named as "rights" or "rights statement". More detailed types of statements may include access rights, terms of use, access condition/access rights, and license.
- Examples of the values (strings or URIs) are from: <http://dublincore.org/usage/meetings/2004/03/dc-rights-proposal.html>

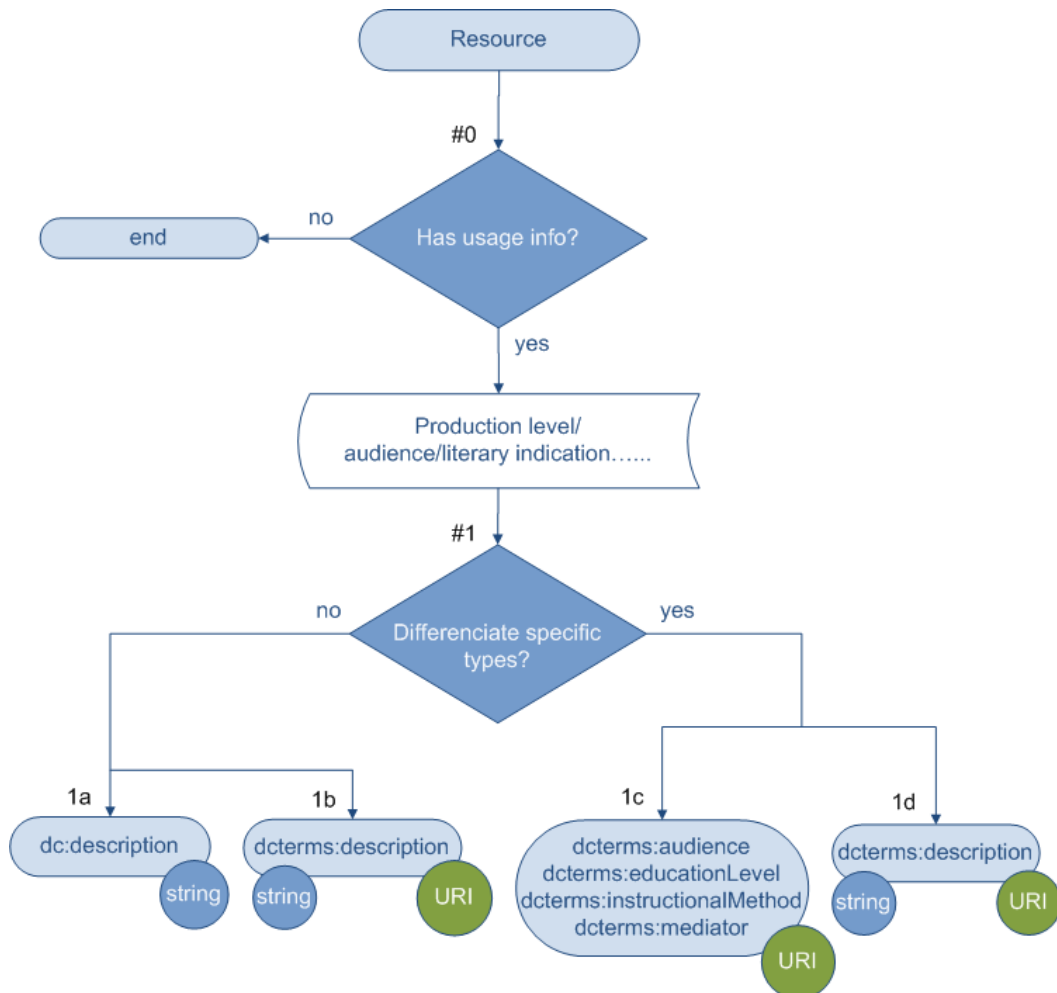
Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Use any controlled vocabulary ?	No	End			
		Yes	Continue to #1			
#1	Differentiate specific types or parts in the rights statement?	No	1a	string	dc:rights	Copyright 1996-2007 XYZ Productions. All rights reserved. http://www.fao.org/corp/copyright/en/
			1b	URI	dcterms:rights	http://www.fao.org/corp/copyright/en/
		Yes	1c	URI	dcterms:rightsHolder	http://www.fao.org/
					dcterms:accessRights	http://www.fao.org/corp/copyright/en/
					dcterms:license	http://creativecommons.org/licenses/by/3.0/
						http://www.fao.org/technical/content/disclaimer-1
					dcterms:rights	[URL of any other type of statements]

3.8. Usage

Properties that are related to the use of a resource, rather than the characteristics of the resource itself, are considered to belong to this group. Typical properties are: *audience*, *literary indication*, and *education Level*.

3.8.1. Audience / literary indication / education Level

Relation with a resource being described: Resource has usage information.



Note

- In presenting the usage-related information, different words might be used in your situation, for example, “Production Level”, “Audience”, “Literary Indication”, etc.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has usage info?	No	End			
		Yes	Continue to #1			
#1	Differentiate specific types of usage data? (e.g., Production level/Audience /Literary indication, etc.)	No	1a	String	dc:description	audience: Public[1]
			1b	String or URI	dcterms:description	audience: Public[1] [URI]
		Yes	1c	URI	dcterms:audience	[rdfs:label "Public"][1]
					dcterms:educationLevel	[rdfs:label "UK Educational Level 1"] [2]
					dcterms:instructionalMethod	[rdfs:label "Direct Teaching"] [3]
					dcterms:mediator	[rdfs:label "Reading specialist"] [4]
					dcterms:description	[String or URI for any other usage data]
				URI or String		

[1] Example taken from ProdiNRA sample record.

[2] Example taken from *UK Educational Levels (UKEL)* list: <http://www.ukoln.ac.uk/metadata/education/ukel/>

[3] Example taken from *ADPRIMA Instructional Methods Information* list of Instructional Methods:<http://www.adprima.com/teachmeth.htm>

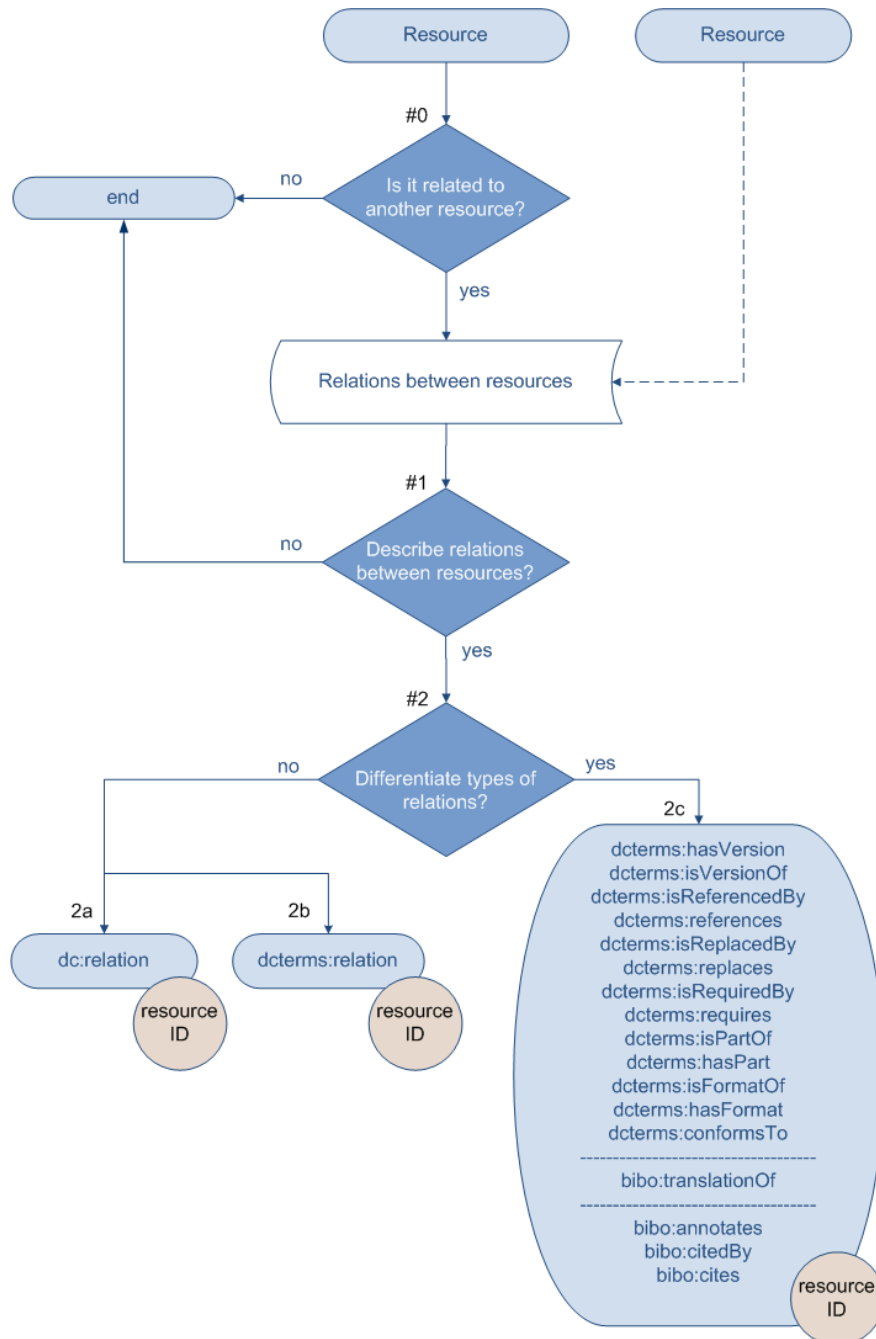
[4] Example taken from Diane Hillmann (2005) *Using Dublin Core*. <http://dublincore.org/documents/usageguide/>

3.9. Relation

This group has a different perspective for describing the resources from other groups that focus on describing the resource itself. Here, various relations *between two resources* or *between two agents* are the focus of the description.

3.9.1. Relation between resources

Relation being described: The resource is related to another resource.



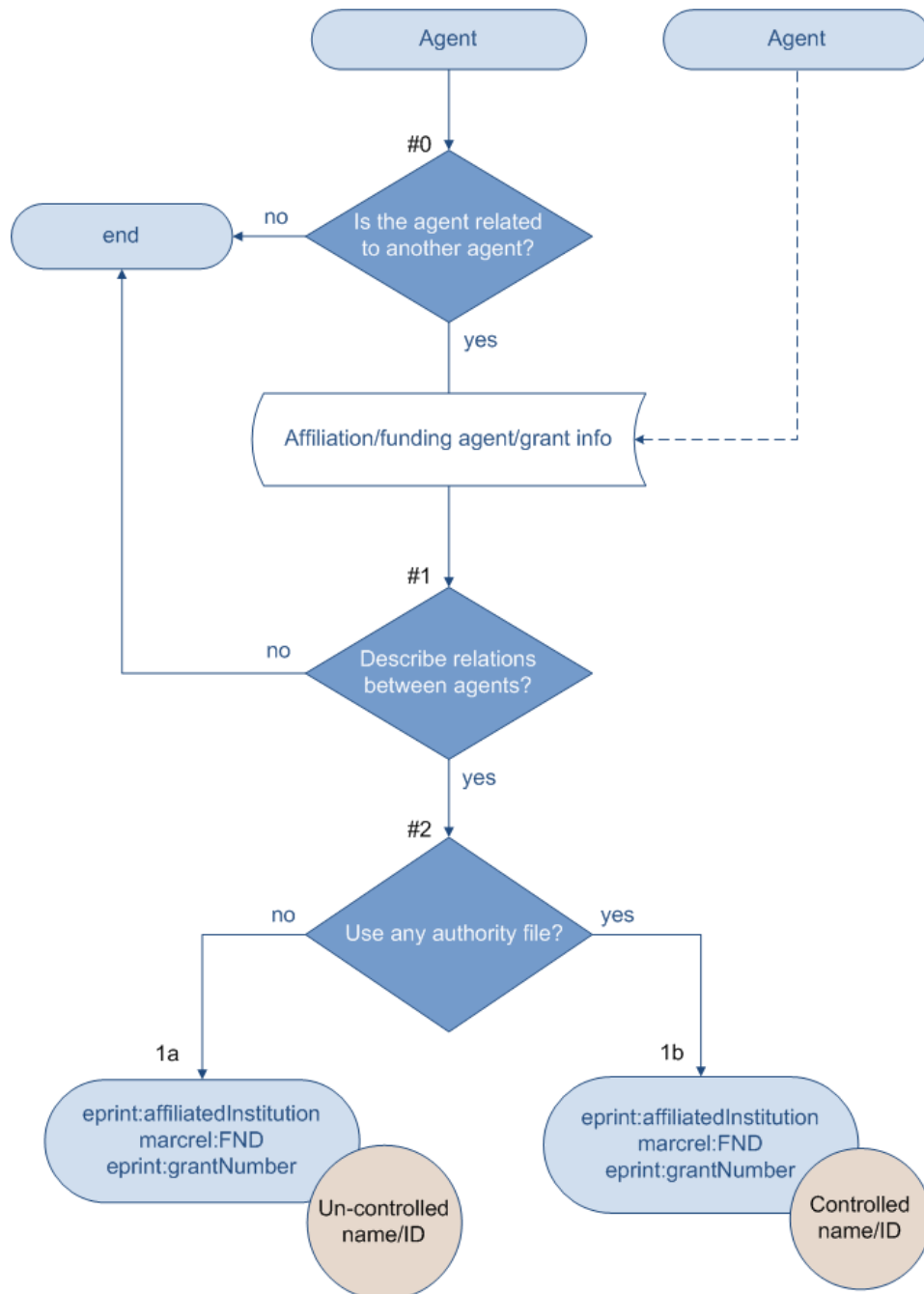
Note

- When a resource is related to another resource, a decision should be made regarding whether the relations between the two resources need to be described.
- In describing the relations, a great number of relation types can be used. The available metadata terms listed below do not form an exhaustive list. Other types may exist.
- The involved resources should always be represented by their identifiers. Values for this property are always the identifiers.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Is it related to another resource?	No	End			
		Yes	Continue to #1			
#1	Describe relations between resources?	No	End			
		Yes	Continue to #2			
#2	Differentiate types of relation?	No	2a	ID	dc:relation	12345
			2b	ID	dcterms:relation	12345
		Yes	2c	ID	dcterms:isVersionOf	12345
					dcterms:hasVersion	
					dcterms:isReplacedBy	
					dcterms:replaces	
					dcterms:isRequiredBy	
					dcterms:requires	
					dcterms:isPartOf	
					dcterms:hasPart	
					dcterms:isReferencedBy	
					dcterms:references	
					bibo:translationOf	
					bibo:annotates	
					bibo:citedBy	
					bibo:cites	

3.9.2. Relation between agents

Relation being described: The agent is related to another agent, specifically affiliation or funding relation.



Note

- When an agent is related to another agent, a decision needs to be made regarding whether the relations between the two agents should be described.
- There could be various types of relations between agents. The available metadata terms listed below focus on the affiliation and funding information and do not form an exhaustive list. Consult *MARC List for Relators* (*marcrel*) <http://id.loc.gov/vocabulary/relators.html> for more types of relators.
- It is highly recommended that agents always be represented by their identifiers or controlled names.

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Is the agent related to another agent?	No	End			
		Yes	Continue to #1			
#1	Describe relations between agents?	No	End			
		Yes	Continue to #2			
#2	Do you use any authority file for the names of the agents?	No	1a	Un-controlled name/ID	eprint:affiliatedInstitution[1]	Univ Bristol
					marcrel:FND [2]	The Mellon Foundation
		Yes	1b	Controlled name/ID	eprint:grantNumber[3]	A456X
					eprint:affiliatedInstitution[1]	University of Bristol
					eprint:grantNumber[3]	A456X
					marcrel:FND [2]	The Andrew W. Mellon Foundation
						http://aims.fao.org/ao/corporate/c_1297 [4]

[1] The eprint:affiliatedInstitution originally has a constrain of domain “ScholarlyWork”, which would not make the use of it for relating agents. EPrint defined “Affiliated Institution” as “An organisation to which a creator of the eprint is affiliated”; LODE-BD considers it appropriate and uses this metadata term without the domain constrain. If a better namespace and metadata term can be identified in the future, LODE-BD will consider a replacement.

[2] marcrel:FND represent “Funder” and has an URI: <http://id.loc.gov/vocabulary/relators/fnd.html>.

– From the *MARC List for Relators*: <http://id.loc.gov/vocabulary/relators/fnd.html>

[3] The eprint:grantNumber originally has a constrain of domain “ScholarlyWork”, which would not make the use of it for relating agents. Based on its definition, a Grant Number is “An alpha-numeric string identifying the funding grant under which the eprint was written”. LODE-BD considers it appropriate and uses this metadata term without the domain constrain. If a better namespace and metadata term can be identified in the future, LODE-BD will consider a replacement.

[4] A corporate body’s URI, from the FAO Authority Description Concept Scheme.

4. The step forward (With Further Readings)

4.1 Implementation Options

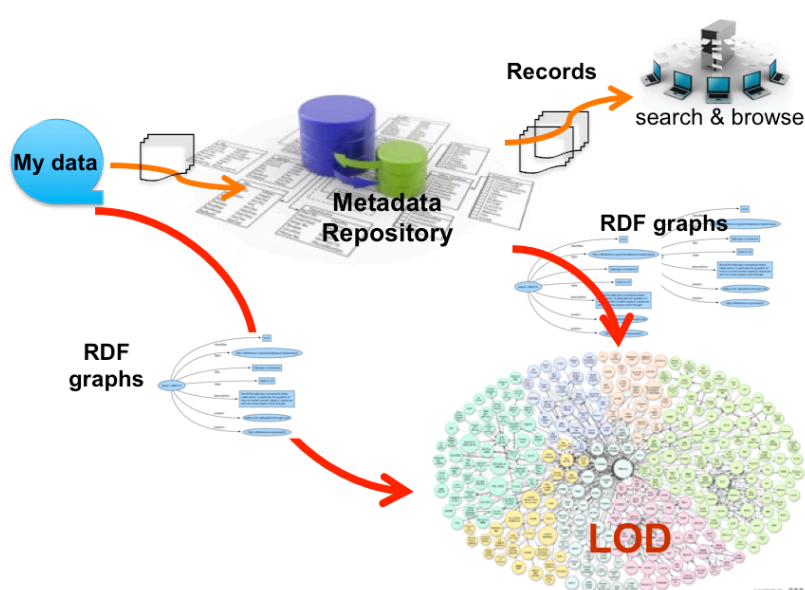
For anyone who is contributing to an open bibliographic data repository and considering preparing LOD-enabled bibliographic data, LODE-BD has provided recommendations on the issues related to processes and strategies. LODE-BD addressed these questions: 1. What kinds of entities and relationships are involved in describing and accessing bibliographic resources? 2. What properties should be considered for publishing meaningful/useful LOD-ready bibliographic data? 3. What metadata terms are appropriate in any given property when producing LOD-ready bibliographic data from a local database? In Section 2, General Recommendations, LODE-BD presents nine groups of common properties identified by LODE-BD and the selected metadata terms to be used for describing bibliographic resources. In Section 3, The Decision Trees, LODE-BD demonstrates how to make decisions on selecting recommended properties according to the local needs.

After metadata terms are selected based on the flowcharts provided in the previous sections, a data provider should have come up with a list of the metadata terms that are appropriate for its existing bibliographic data. To implement these metadata terms, LODE-BD shares two options that have been summarized based on the best practices.

- Option #1, "Design-time" strategy: The data provider would need to change its current ad-hoc model, replacing it with the LODE-BD "good practices" model and those selected metadata terms. This means some changes to a database and the services that access it.
- Option #2, "Run-time" strategy: The data provider would keep the original ad-hoc model and database structure. A data provider would convert bibliographic data on the fly to a "good-practices" model upon request. This means to add a conversion service and leave the ad-hoc model unchanged.

Turning the bibliographic data from an ad-hoc modeled database in a silo to the data in a standardized metadata repository, it is a giant leap because the unified data records from various data providers can be maximized in searching and browsing through the services of the repository. Furthermore, the same practice could also lead to a step heading to the LOD universe. The individual data providers can directly produce RDF triples using LODE-BD recommended metadata terms. Or, this mission can be accomplished through the metadata repository, which would publish its bibliographic data as Linked Data, as illustrated in the following figure. In both outcomes, preparing LOD-ready metadata by data providers is essential.

Figure 2. Output of LOD-Ready Metadata



4.2 How to publish and consume Linked Data

Linked Data: Evolving the Web into a Global Data Space (1st edition), Tom Heath and Christian Bizer (2011). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

URL: <http://linkeddatabook.com>

"This book gives an overview of the principles of Linked Data as well as the Web of Data that has emerged through the application of these principles. The book discusses patterns for publishing Linked Data, describes deployed Linked Data applications and examines their architecture."

Linked Data Patterns, Leigh Dodds and Ian Davis. (2011).

URL: <http://patterns.dataincubator.org/book/>

"A pattern catalogue for modelling, publishing, and consuming Linked Data."

Linked Data star scheme by example

URL: <http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>

"Tim Berners-Lee suggested a 5-star deployment scheme for Linked Open Data and Ed Summers provided a [nice rendering](#) of it."

Linked Data - Design Issues, Tim Berners-Lee (2006).

URL: <http://www.w3.org/DesignIssues/LinkedData.html>

One of the first discussions of the topic, mentioning the *"four rules of Linked Data"*.

Cool URIs for the Semantic Web. Leo Sauermann and Richard Cyganiak (2008). W3C Interest Group Note.

URL: <http://www.w3.org/TR/cooluris/>

4.3. Where to find Linked Data sets and Vocabularies

CKAN Data Hub

CKAN is a metadata registry for datasets. Many of the datasets described in CKAN are in linked-data form. The datasets are described by curators regarding their dataset size, example resources and access methods (e.g. SPARQL endpoints) and, crucially, links to other datasets.

Linked Open Data Cloud

URL: <http://ckan.net/group/lodcloud>

Datasets in the Linking Open Data (LOD) Cloud diagram. It is based on metadata collected and curated by contributors to the [CKAN directory](#). Each dataset is a hyperlinked from the diagram to its homepage.

The Linking Open Data cloud diagram.

URL: <http://richard.cyganiak.de/2007/10/lod/>

Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets, W3C Incubator Group Report 25 October 2011

URL: <http://www.w3.org/2005/Incubator/llid/XGR-llid-vocabdataset-20111025/>

A side delivery of the W3C Linked Library Data (LLD) XG which lists relevant metadata element sets, value vocabularies that are reported in the Linked Library Data use cases and case studies. Each entry contains link URL, namespace, and short description.

Linked Open Vocabularies (LOV)

URL: <http://labs.mondeca.com/dataset/lov/index.html>

A dataset of descriptions of RDFS vocabularies or OWL ontologies defined for and used by LD datasets.

4.4. How to express metadata with different syntaxes: text, html, xml, rdf, and rdfa

DC-TEXT [DCMI Recommendation]. "Expressing Dublin Core metadata using the DC-Text format"

URL: <http://www.dublincore.org/documents/dc-text/>

Its primary use is in presenting metadata constructs for human consumption.

DC-HTML [DCMI Recommendation]. "Expressing Dublin Core metadata using HTML/XHTML meta and link elements"

URL: <http://www.dublincore.org/documents/dc-html/>

It describes how a Dublin Core metadata description set can be encoded using the HTML/XHTML <meta> and <link> elements. This specification is also an HTML "meta data profile" as defined by the HTML specification.

DC-DS-XML [DCMI Proposed Recommendation]. "Expressing Dublin Core Description Sets using XML (DC-DS-XML)"

URL: <http://www.dublincore.org/documents/dc-ds-xml/>

It specifies an XML format for representing a Dublin Core metadata description set.

DC-RDF [DCMI Recommendation]. "Expressing Dublin Core metadata using the Resource Description Framework (RDF)"

URL: <http://www.dublincore.org/documents/dc-rdf/>

It describes how constructs of the DCMI Abstract Model may be expressed in RDF graphs.

User Guide/ Publishing Metadata.

URL: http://wiki.dublincore.org/index.php/User_Guide/Publishing_Metadata

"How to use DCMI Metadata as linked data."

Linked Data Tutorial NG - Publishing and Consuming Linked Data with RDFa, Michael Hausenblas and Richard Cyganiak.

URL: <http://ld2sd.deri.org/lod-ng-tutorial/>

"This note describes, step-by-step, how to create and consume linked data with RDFa."

4.5. Why publish bibliographic data as Linked Data?

Library Linked Data Incubator Group Final Report, W3C Incubator Group Report 25 October 2011.

URL: <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>

- Benefits of the Linked Data Approach
http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/#Benefits_of_the_Linked_Data_Approach
- Recommendations
<http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/#Recommendations>

Acknowledgement

This version is based on LODE-BD 1.1 that was partially supported by the European Commission through the ICT PSP Grant #250525 for VOA3R (Virtual Open Access Agriculture & Aquaculture Repository: Sharing Scientific and Scholarly Research related to Agriculture, Food, and Environment). The authors also would like to thank Ioannis N. Athanasiadis, Nikos Manouselis, Ilias Hatzakis, Tom Baker, Gordon Dunsire, Hugo Besemer, Fernanda Peset, Xavier Agenjo, Francisca Hernández, MacKenzie Smith, Karen Coyle, Antoine Issac, the FAO AIMS Group, and the data providers of the VOA3R team for their support and advice throughout the completion of this project, as well as Laurence Skirvin who assisted in the compiling of the crosswalk to schema.org in the added appendix 4. The authors also want to thank Gordon Dunsire especially for his continuing advising on the current LODE-BD 2.0 version.

Appendixes

Appendix 1. Background of the original report, LODE-BD 1.1 Encoding bibliographic data within the VOA3R Federation and Beyond

The idea of assisting information professionals in deciding what metadata terms to use when encoding existing bibliographic data for the purpose of exchanging and sharing across data providers was born under the umbrella of [VOA3R](http://voa3r.eu/), an European research consortium project. VOA3R stands for *Virtual Open Access Agriculture & Aquaculture Repository: Sharing Scientific and Scholarly Research related to Agriculture, Food, and Environment*.⁴ The general objective of the VOA3R project is to improve the spread of European agriculture and aquaculture research results by using an innovative approach to sharing open access research products. Under a strict open access policy, the VOA3R Federation connects libraries, archives and other publication systems by providing advanced search interfaces that include specifics aspects of research work (methods, variables, measures, instruments, techniques, etc.) of each particular domain. The users of the VOA3R service are not only researchers, but also students and practitioners who want to either search for or publish scientific research results. The project is targeted to the domain of agriculture & aquaculture, as it re-uses previous models for these domains, but the technology and models integrated are largely transferable to other academic disciplines and subject domains.

The VOA3R Federation is composed of 17 institutions from 13 countries which contribute bibliographic data to eight open repositories. In order to exchange metadata, VOA3R originally planned to use two different application profiles. The first one would be a VOA3R Application Profile based on the Dublin Core Metadata Element Set (DC); The second one would be an application profile based on the Metadata Object Description Schema (MODS), which would be used by those repositories that have richer bibliographic data.

After a series of discussions within the VOA3R Federation, a new idea regarding the data harvesting approach emerged. In addition to the original plan, a set of recommendations was foreseen with a full range of options for metadata encoding which data providers could choose from according to their development stages, internal data structures, and the reality of their practices. Through these recommendations, the VOA3R data providers should get simple answers for questions like: “Will the physical holding of a resource be important enough to be shared among the VOA3R participants?” “What metadata term should be used for encoding the title(s), identifier(s), or subject(s)?”

It was also decided that the recommendations would allow any data provider to encode bibliographic data using properties from standardized namespaces, to use well-established authority data and controlled vocabularies that are available as linked data in agriculture and aquaculture, to publish data in RDF triples, and to submit the dataset to VOA3R. In doing so, VOA3R would act both as a service provider enhancing the dissemination channel and accessibility of open access documents and as a service that promotes the exchange and publication of bibliographic data in RDF, so as to facilitate the use of Linked Data in agriculture and aquaculture.

In this context, the LODE-BD Recommendations were prepared. In spite of the fact that the recommendations are geared toward the agriculture and aquaculture sectors through the VOA3R project, the recommendations are destined to become useful for any type of bibliographical data describing bibliographic resources in any subject domain.

⁴ VOA3R <http://voa3r.eu/>

Appendix 2. Explanation of Terminology

Certain terminology has been applied throughout this report. Short explanations are provided below.

Metadata Terms and Properties

“[metadata] elements”, “[metadata] fields”, and “attributes [of an entity]” have been widely used by the professionals who are involved in creating, designing, and implementing metadata standards. In a number of metadata structure standards it is the term “elements” that have dominated in the specifications. Some standards (e.g., those used by library, museum, and archives communities) prepared their data structure standards (e.g., MODS, CDWA, VRA Core 4, EAD) using XML schema as the primary medium. These specifications modelled the structure with a set of “elements” and sub-elements, related “attributes”, and controlled “attribute values” throughout the element sets. Nevertheless, as represented by DCMI Metadata Terms (DCTERMS), the RDF terminology instead of the XML terminology is now gaining momentum. The term “properties” of resources are used in place of “elements” in this report. LODE-BD considers the process of metadata description as the description of properties of a resource. For example, ‘rights’ is considered as a property of a resource.

Property: rights

Because there are various levels of granularity and several corresponding ways that this property can be described, LODE-BD uses “*metadata term*” for a specific element formally defined by a metadata namespace. For example, property ‘rights’ can be described by metadata terms from different namespaces:

Metadata term: dc:rights

Metadata term: dcterms:rights

String and URI as values

In the LODE-BD Recommendations, the words ‘string’ and ‘URI’ are used for the most commonly seen values in bibliographic data. They correspond to the terminology of RDF in the form of ‘literal’ (typically a string of characters) and ‘non-literal’.

Literal: “The most primitive value type represented in RDF, typically a string of characters. The content of a literal is not interpreted by RDF itself and may contain additional XML markup. Literals are distinguished from Resources in that the RDF model does not permit literals to be the subject of a statement.”⁵

Non-literal value: “A value which is a physical, digital or conceptual entity.”⁶

For example, “rice” is a concept included in the *AGROVOC Thesaurus*, with a preferred label (in English), “Rice.” When the thesaurus is published as Linked Data, the concept is considered as a resource and is given a unique URI, http://aims.fao.org/aos/agrovoc/c_6599. This means that a URI reference is used to identify this concept as a resource.

In this situation for the *property*: subject, the *metadata terms* for encoding this property include *dc:subject* and *dcterms:subject*. Because *dcterms:subject* “is intended to be used with non-literal values as defined in the DCMI Abstract Model (<http://dublincore.org/documents/abstract-model/>)”,⁷ the value to be used associated with this metadata term should be the URI http://aims.fao.org/aos/agrovoc/c_6599 which represents the concept as a resource instead of “Rice” or other language labels of the concept.

Based on the definition of these metadata terms, the following examples are provided:

Metadata Term	Value Type	Example
dc:subject	String	Rice
dcterms:subject	URI	http://aims.fao.org/aos/agrovoc/c_6599

[Bibliographic] Resource

The term “Resource” is used in the conceptual model to denote a general entity, the *Bibliographic Resource*. An instance of the bibliographic resource can be an article, monograph, thesis, conference paper, research report, presentation material, learning object, etc., regardless if it is in print or electronic format. In the flowcharts provided by the LODE-BD Recommendations, the ‘resource’ at the beginning oval box is an instance of the Bibliographic Resource.

⁵ Resource Description Framework (RDF) Model and Syntax Specification (1999-02-22). Glossary for this source <http://www.w3.org/2003/glossary/keyword/All/literal.html?keywords=literal> Last accessed February 2011

⁶ DCMI Abstract Model. <http://www.dublincore.org/documents/abstract-model/> Last accessed February 2011

⁷ DC Terms. <http://dublincore.org/documents/dcmi-terms/#terms-subject> Last accessed May 28, 2012.



Appendix 3. Metadata Standards used in LODE-BD

A selected number of widely-used metadata standards and the emerging LOD-enabled vocabularies for bibliographic descriptions are used as the basis for the metadata terms recommended in LODE-BD.

dc

Dublin Core Metadata Element Set (DCMES or DC)

Dublin Core Metadata Initiative (DCMI)

Namespace: <http://purl.org/dc/elements/1.1/>

Page: <http://dublincore.org/documents/dces/>

The *Dublin Core Metadata Element Set* (ISO 15836) is a vocabulary of fifteen properties for use in resource description.

dcterms

DCMI Metadata Terms

Dublin Core Metadata Initiative (DCMI)

Namespace: <http://purl.org/dc/terms/>

Page: <http://dublincore.org/documents/dcmi-terms/>

The *DCMI Metadata Terms* is an authoritative specification of all metadata terms maintained by DCMI. As a full set of DCMI vocabularies it also includes sets of resource classes (including the DCMI Type Vocabulary), vocabulary encoding schemes, and syntax encoding schemes.

bibo

Bibliographic Ontology

Bibliographic Ontology Specification Group

Namespace: <http://purl.org/ontology/bibo/>

Page: <http://bibliontology.com/specification>

The *Bibliographic Ontology* is designed for use in describing bibliographic things on the semantic Web in RDF.

agls

AGLS Metadata Standard

Australian Government Locator Service

Namespace: <http://www.agls.gov.au/agls/terms/>

Page: <http://www.agls.gov.au/documents/aglsterms/>

The *AGLS Metadata Standard* (Australian Standard AS 5044-2010) is developed to promote consistency of discovery of government resources. It provides a set of metadata properties and associated usage guidelines to improve the visibility, manageability and interoperability of online information and services.

eprint

Eprints Terms

UKOLN, JISC

Namespace: <http://purl.org/eprint/terms/> Page: Page:

http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Terms

The *Eprints Terms* include eprints-specific metadata properties and encoding schemes that have been created as part of the Dublin Core-based *Scholarly Works Application Profile*.

marcrel

MARC List for Relators

Library of Congress

Namespace: <http://id.loc.gov/vocabulary/relators/>

Page: <http://www.loc.gov/marc/relators/relators.html>

Relator terms and their associated codes are originally designed for use with the MARC records, for designating the relationship between a name and a bibliographic resource.

4. Holding/Location Information			bibo:eanucc13		
			bibo:eissn		
			bibo:gtin14		
			bibo:handle		
			bibo:isbn	isbn	. . Book
			bibo:issn		
			bibo:iccn		
			bibo:oclcnum		
			bibo:pmid		
			bibo:sici		
			bibo:upc		
			bibo:uri		
	dc:language	dcterms:language		Inlanguage	. CreativeWork
	dc:format	dcterms:format	dcterms:medium	> bookEdition	. . Book
	bibo:edition			> version	. CreativeWork
	bibo:status				
	dc:source	dcterms:source	bibo:pages	> numberOfPage	. . Book
			bibo:pageStart		
			bibo:section		
			bibo:volume		
			bibo:issue		
			bibo:chapter		
			bibo:locator	contentLocation	. CreativeWork
	agls:availability				

5. Subject Information	dc:subject	dcterms:subject		about	. CreativeWork
	dc:coverage	dcterms:coverage	dcterms:spatial	keywords	. CreativeWork
6. Description of Content			dcterms:temporal	mentions	. CreativeWork
	dc:description	dcterms:description	dcterms:abstract	keywords	. CreativeWork
7. Intellectual Property Rights	dc:type	dcterms:type	dcterms:tableOfContent	description	Thing
	dc:rights	dcterms:rights	dcterms:rightsHolder	genre	. CreativeWork
			dcterms:accessRights	copyrightHolder	. CreativeWork
			dcterms:license		
8. Usage	dc:description	dcterms:description		description	. CreativeWork
		dcterms:audience	dcterms:educationLevel	audience	. CreativeWork
9. Relation [between resources]		dcterms:instructionalMethod	dcterms:mediator		
			dcterms:isVersionOf		
	dc:relation	dcterms:relation	dcterms:hasVersion		
			dcterms:isReplacedBy		
			dcterms:replaces		
			dcterms:isRequiredBy		
			dcterms:requires		
			dcterms:isPartOf		
			dcterms:hasPart		
			dcterms:isReferencedBy		
[between agents]			dcterms:references		
			bibo:translationOf		
			bibo:annotates		
			bibo:citedBy bibo:cites		
	eprint:affiliatedInstitution		affiliation		. Person
	eprint:grantNumber		>sourceOrganization		. CreativeWork
	marcel:FND		> sponsor		. MedicalEntity . . MedicalStudy